

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 22, 2002, 17:51:15 ; Search time 79.11 Seconds
(without alignments)
625.470 Million cell updates/sec

```
Title: US-09-265-585C-96
Perfect score: 3503
Sequence: 1 MPPPPPPPELTPYCRRCPP.....KDLCLLTASAMRP IQVPPCR 668
```

Scoring table:	BLOSUM62	Ganex	0.5
	Ganex	10.0	Ganex

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1: /SIDS2/gcgcata/genseq/genseqp/AA1960.DAT: *

A_genseq_1101.*

2: /SIDS2/gcgcata/genseq/genseqp/AA1981.DAT: *

3: /SIDS2/gcgcata/genseq/genseqp/AA1982.DAT: *

4: /SIDS2/gcgcata/genseq/genseqp/AA1983.DAT: *

5: /SIDS2/gcgcata/genseq/genseqp/AA1984.DAT: *

6: /SIDS2/gcgcata/genseq/genseqp/AA1985.DAT: *

7: /SIDS2/gcgcata/genseq/genseqp/AA1986.DAT: *

8: /SIDS2/gcgcata/genseq/genseqp/AA1987.DAT: *

9: /SIDS2/gcgcata/genseq/genseqp/AA1988.DAT: *

10: /SIDS2/gcgcata/genseq/genseqp/AA1989.DAT: *

11: /SIDS2/gcgcata/genseq/genseqp/AA1990.DAT: *

12: /SIDS2/gcgcata/genseq/genseqp/AA1991.DAT: *

13: /SIDS2/gcgcata/genseq/genseqp/AA1992.DAT: *

14: /SIDS2/gcgcata/genseq/genseqp/AA1993.DAT: *

15: /SIDS2/gcgcata/genseq/genseqp/AA1994.DAT: *

16: /SIDS2/gcgcata/genseq/genseqp/AA1995.DAT: *

17: /SIDS2/gcgcata/genseq/genseqp/AA1996.DAT: *

18: /SIDS2/gcgcata/genseq/genseqp/AA1997.DAT: *

19: /SIDS2/gcgcata/genseq/genseqp/AA1998.DAT: *

20: /SIDS2/gcgcata/genseq/genseqp/AA1999.DAT: *

21: /SIDS2/gcgcata/genseq/genseqp/AA2000.DAT: *

22: /SIDS2/gcgcata/genseq/genseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3503	100.0	668	21	AAB28595	Maize SCARECROW protein encoded by SCARECROW1
2	1861	53.1	653	18	AAM38178	Arabidopsis SCARECROW1 protein encoded by SCARECROW1
3	1861	53.1	653	21	AAB28596	Arabidopsis SCARECROW1 protein encoded by SCARECROW1
4	1131.5	32.3	716	18	AAM38192	Maize ZCARECROW2 protein encoded by ZCARECROW2
5	1131.5	32.3	716	21	AAB28591	Maize ZCARECROW2 protein encoded by ZCARECROW2
6	1005.5	28.7	405	21	AAB28516	Maize ZCARECROW2 protein encoded by ZCARECROW2
7	954.5	27.2	323	21	AAB28603	Arabidopsis thaliana SCARECROW1 protein encoded by SCARECROW1
8	705.5	20.1	246	21	AAAC5517	Maize CBPRT44 protein encoded by CBPRT44
9	670	19.1	623	20	AAV02540	Arabidopsis thaliana SCARECROW1 protein encoded by SCARECROW1
10	666	19.0	630	20	AAV02541	Arabidopsis thaliana SCARECROW1 protein encoded by SCARECROW1
11	651	18.6	240	21	AAG23966	Arabidopsis thaliana SCARECROW1 protein encoded by SCARECROW1

12	6.18	17.6	425	20	AAV02559	Protein encoded by
13	594.5	17.0	587	22	AAE02560	<i>A. thaliana</i> transp
14	594.5	17.0	587	22	AAE01892	<i>Arabidopsis thalia</i>
15	590.5	16.9	517	21	AAE03857	<i>Arabidopsis thalia</i>
16	590.5	16.9	518	21	AAE038576	<i>Arabidopsis thalia</i>
17	590.5	16.9	532	18	AAAG0792	<i>Arabidopsis thalia</i>
18	590.5	16.9	533	21	AAE038575	<i>Arabidopsis thalia</i>
19	589.5	16.8	531	21	AAE038575	<i>Arabidopsis thalia</i>
20	589.5	16.8	532	22	AAE02545	<i>Arabidopsis thalia</i>
21	589.5	16.8	532	22	AAE01907	<i>Arabidopsis thalia</i>
22	589.5	16.8	572	22	AAE01883	Amino acid sequenc
23	585.5	16.7	572	22	AAE01884	Amio acid sequenc
24	548	15.6	384	21	AAE029710	<i>Arabidopsis thalia</i>
25	548	15.6	405	21	AAE029709	<i>Arabidopsis thalia</i>
26	548	15.6	413	21	AAE029708	<i>Arabidopsis thalia</i>
27	544.5	15.5	480	21	AAE01049	<i>Arabidopsis thalia</i>
28	544.5	15.5	578	21	AAE01048	<i>Arabidopsis thalia</i>
29	544.5	15.5	593	21	AAE01047	<i>Arabidopsis thalia</i>
30	532.5	15.2	587	21	AAE08574	<i>Arabidopsis thalia</i>
31	525	15.0	428	19	AAE01753	Tomato Ls protein.
32	502.5	14.3	306	21	AAE08558	Protein encoded by
33	499.5	14.3	306	18	AAE08179	<i>Arabidopsis thalia</i>
34	499.5	14.3	306	21	AAE08570	<i>Arabidopsis thalia</i>
35	498	14.2	352	21	AAE08556	Protein encoded by
36	492	14.0	352	18	AAE08181	<i>Arabidopsis thalia</i>
37	466	13.3	541	21	AAE00885	<i>Arabidopsis thalia</i>
38	466	13.3	543	21	AAE00884	<i>Arabidopsis thalia</i>
39	466	13.3	510	21	AAE00883	<i>Arabidopsis thalia</i>
40	439.5	12.5	513	21	AAE01536	<i>Arabidopsis thalia</i>
41	439.5	12.5	515	21	AAE01535	<i>Arabidopsis thalia</i>
42	439.5	12.5	583	21	AAE01534	<i>Arabidopsis thalia</i>
43	423.5	12.1	375	21	AAE02546	<i>Arabidopsis thalia</i>
44	404.5	11.5	287	21	AAE02547	<i>Arabidopsis thalia</i>
45	404.5	11.5	809	21	AAE02546	<i>Arabidopsis thalia</i>

ALIGNMENTS

RESULT	1
AA28595	
ID	AA28595 standard; Protein; 668 AA.
XX	
AC	AA28595;
XX	
DT	09-FEB-2001 (first entry)
DE	Maize Scarecrow protein.
XX	
KW	Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
KW	transgenic plant; cell division; molecular marker; herbicide resistance
KW	salt resistance; pathogen resistance; insect resistance.
OS	2ea mays.
XX	
PN	W0200053723-A2.
PD	14-SEP-2000.
XX	
PF	07-MAR-2000; 2000MO-US05875.
XX	
PR	10-MAR-1999; 99US-0265585.
XX	
PA	(UNYNY) UNIV NEW YORK STATE.
XX	
PI	Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L;
PI	Helariutta Y, Bruce W, Läm J;
XX	
DR	WPI: 2000-594315/56.
DR	N-PSDB: AAC65299.
XX	
PT	Scarecrow gene useful for producing transgenic plants expressing genes
PT	whose product increases starch, lignin or cellulose biosynthesis and

PT confers herbicide, pathogen or insect resistance -
 XX
 PS Claim 12; Fig 25; 200pp; English.
 XX
 CC The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MOTIF III (VHID) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryos and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.
 CC
 XX
 SQ Sequence 668 AA;

Query Match 100.0%; Score 3503; DB 21; Length 668;
 Best Local Similarity 100.0%; Pred. No. 5,8e-269;
 Matches 668; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPPPPPPPLPYCRCPRHLP PPPSSPNHLLHYHLDHQAAMAAVRRKRPASDM 60
 DB 1 MPPPPPPPLPYCRCPRHLP PPPSSPNHLLHYHLDHQAAMAAVRRKRPASDM 60
 QY 61 DLPRRRHVTGDLSDVTAAGVSGAPSSASQALPALTQQLPAPFOHNAPEVDY 120
 DB 61 dlprrrhvtgdlsvtaagvsgapsasqalpaltqlhqlpafghnapevdv 120
 QY 121 PNHPRAAHQAQGEATSTTAWVDGTRDITIGSSGAASVTQLIHNVRELIHPCNGL 180
 DB 121 pnhpraaahqaqgeatsttawvdlrldlgsagsaavsltqlihnvrelihpcnpgl 180
 QY 121 PNHPRAAHQAQGEATSTTAWVDGTRDITIGSSGAASVTQLIHNVRELIHPCNGL 180
 DB 121 pnhpraaahqaqgeatsttawvdlrldlgsagsaavsltqlihnvrelihpcnpgl 180
 QY 181 ASLLELRSLAADPAPLP PPPPOFOHALHGAFAAPAGTLPPRPLPDKRHEHPP 240
 DB 181 asllelrslaaadpaplpppppqghallhgapaapagltlpppplpdkrthennpp 240
 QY 241 PCQOQOQOEEHPAPQSPKAPTAETAAAAAQAAMAAARERKEERKORODEGLTL 300
 DB 241 pcqggqgeehpapgspkaptaeetaaaaaqaamaaakerkeeqgrtkrdeeghl 300
 QY 241 PCQOQOQOEEHPAPQSPKAPTAETAAAAAQAAMAAARERKEERKORODEGLTL 300
 DB 241 pcqggqgeehpapgspkaptaeetaaaaaqaamaaakerkeeqgrtkrdeeghl 300
 QY 301 TLLLOCAKAVNANDNDDAHQTLLLETAELATPGTGSTORVAAVPAFAMSAKLVSSCLGIXA 360
 DB 301 tlllqcaevnandnddahqtllelaelatlpgtstgrvaayfaaamsarlvsclgilya 360
 QY 361 PLPPGSPAARLHGKVAFAQVFNCSIPVFKEFHPTANQAOIOEAFEREERHIIIDLDIMO 420
 DB 361 plppgsaaarlhgkvaafqvfngcspfvkfshftanqaiqeafereerhiidldimg 420
 QY 421 GLQWEGLFHIIASRGGPPRVRLTGLGASMEALATGRKLSDFADTLGLPEFCVAEKA 480
 DB 421 glqweglfhiiasrpgpprvrltglgasmaealeatgrklsdfadtlglpfcfaeaka 480
 QY 481 GNWDEKLGVTFRREAVAVHMLHSLYDVTGSDSNLMLIQRLAKRVYVMWDLSHSGSF 540
 DB 481 gnwdekigvtfrreavavhmlhsltydvtgdsnltlwiqlapkvvmwvgedlshsgsf 540
 QY 541 LARFEALHYYSALFDSDASYGEDSPRHVVEOQLSREIRNVLAVGAPARTGDVKGFS 600
 DB 541 larfealhyysalfdsdasygedsprhvveqqlsreirnvlaavgpartgdvkgfs 600
 QY 601 WRKLAQSGFRAASLAGSAAAQASLLQMFPSDGYTTLVEENGAKLKGKDCILTLTASMR 660
 DB 601 wrklaqsgftraaslagasaaqasllqgmfsdgytlyveengalklgykdcil tlaaswr 660
 QY 661 PIQVPPCR 668

DB 661 piqvppcr 668
 RESULT 2
 AAW38178
 ID AAW38178 standard; Protein; 653 AA.
 XX
 AC AAW38178;
 XX
 DT 11-MAY-1998 (first entry)
 XX
 DE Arabidopsis SCARECROW protein.
 XX
 KW SCARECROW; SCR gene; transgenic plant; root; gravitropism;
 XX crop improvement.
 OS Arabidopsis thaliana.
 XX
 FH Key
 FH Location/Qualifiers
 FT Region
 FT 1..267
 FT /note= "N-terminal region rich in Gln, Ser, Thr and
 FT Pro homopolymeric structures"
 FT Peptide
 FT 291..322
 FT /note= "Ieucine heptad repeat"
 FT Peptide
 FT 436..473
 FT /note= "Ieucine heptad repeat"
 FT Domain
 FT 58..99
 FT /label= MOTIF-I
 FT Domain
 FT 265..341
 FT /label= MOTIF-II
 FT Domain
 FT 373..435
 FT /label= MOTIF-III (VHID)
 FT Domain
 FT 436..480
 FT /label= MOTIF-IV
 FT Domain
 FT 457..496
 FT /label= MOTIF-V
 FT Domain
 FT 516..649
 FT /label= MOTIF-VI
 PN W09741152-A1.
 XX
 PD 06-NOV-1997.
 XX
 PF 25-APR-1997; 97WC-US07022.
 XX
 PR 24-APR-1997; 97US-0842445.
 PR 26-APR-1996; 96US-0638617.
 XX
 PA (UNY) UNIV NEW YORK STATE.
 XX
 PI Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
 PI Pish L, Mysiocka-Diller J;
 PI
 DR WPI; 1997-549683/50.
 DR N-PSDB; AAT95753.
 XX
 PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
 PT cell division and therefore alter root development, or alter plant
 PT stem or hypocotyl gravitropism
 PS
 PS Claim 12; Page 92-94; 221pp; English.
 CC This is the deduced amino acid sequence of the Arabidopsis
 CC SCARECROW (SCR) gene (see AAT95753). SCR is a member of a novel
 CC protein family and contains a number of potential functional
 CC domains similar to those found in transcription factors. SCR is
 CC expressed specifically in embryo root progenitor tissue and in
 CC certain root and stem tissues. It regulates a specific asymmetric
 CC division, and controls gravitropic response in aerial structures
 CC and root formation. SCR proteins (see AAW38178-201) from dicots
 CC monocots, SCR proteins lacking 1-4 of MOTIFS I-VI, and polypeptides
 CC corresponding to MOTIF I, II, III, IV, V or VI of SCR are claimed.

CC Transgenic plants can be engineered to overexpress the SCR protein,
 CC so that cell division is increased in roots, resulting in thicker
 CC root development, while a plant with an altered stem or hypocotyl
 CC gravitropism is less susceptible to lodging. Plants that contain
 CC an antisense molecule that suppresses the expression of endogenous
 CC SCR gene product have thinner root development. A gene of interest
 CC can be placed under control of a SCR promoter and expressed in a
 CC plant to confer herbicide, salt, pathogen or insect resistance, or
 CC when expressed in stems to increase starch, lignin or cellulose
 CC biosynthesis (all claimed).

Sequence 653 AA;

Query Match 53.1%; Score 1861; DB 18; Length 653;
 Best Local Similarity 57.3%; Pred. No. 6.2e-139;
 Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

QY 4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHLHLDHQAAMAAWVKRP 56
 DB 12 PPPHSPILRTSSGSSSSNNRPPP--PPPP-----LVMVTKRL 48
 QY 57 ASDM-----DLPPRR--H-VTGDLSVT-----AAAAGVGGSGAPS---SASAQ 96
 DB 49 asemssnpdymnsrrprvshllidsnyntvtpqgppltaaatvssqpnplsvcgfsg 108
 QY 97 LPLPTQLHQLPRAFQNHAPVNDVPAHPAPAAHQAAGEATASTTAWDGIIRDIIGSSG 156
 DB 109 lpfvfrs-----drggrtnvmsvqpmnd---qdssssasptvwdaifrdllhss- 154
 QY 157 GAAVSITQLIHNVRELIHPCNPGLASILELRSLAADPA-----PLP----- 200
 DB 155 -tsvisipqlignvrdlflfcpnplgallelyrltsimldpsssdsbpqtfelyqismn 213
 QY 201 --PPPOQHALLHGAPAAAPAGTLPPPPPLDPRKRHEHPPQOOQOOEHPHAPSPK 258
 DB 214 pspqgqgqbq-----qgqgqbhkprrppidqgere-----nstd 248
 QY 259 APTRAEETAAAA-AAQAAAAAAKERKEEQRKORDEBGLHLLTLQLCAEAVNADNLD 317
 DB 249 appqetvlatvpavqntaetalkrkeekrkqdeegllhlllllqcaevadnlee 308
 QY 318 AHQTLLELAELATPGTSTORVAVFAEAMASRLVSSCIGIYALPFGSPRAAALHG-RV 376
 DB 309 anklllelsgdstpygsaqvvaayfseamaarlinsclgylaLP--strwmpqthslkm 366
 QY 377 AAARQVFNIGSPFYKFSHFTANQAIQEAFFEREERVHIIIDIMQGLQMPGLFHLIASRP 436
 DB 367 vsatqvfngisplvkrfshtfngaiqeaefekedsvhlldimngliqwpqlfhljastrpy 426
 QY 437 GPPRYRLTGLGASMEALEATGKRLSDPADTLGLPFECAVAEKAQNVDPKLGVTREAV 496
 DB 427 gpphvrllgltgtsmeaiaqatgkrlsdfdklglrpfefcplaekvgnldterlnvrkrea 486
 QY 497 AVHMHLSLTVNVTGSDSTWTLILQRLAPKYVTWYEOQLSHSGSLAFVFAIHYSALFD 556
 DB 487 avhwqlshlyvngsdadtlwlgtrlapkvvtlvveqdlshagsrlgrtveahlyysalfd 546
 QY 557 SLDASGDESPERHVEEQQLSREIRNVLAVGAPARTGDVFGSMRKLAKOSGRASLAS 616
 DB 547 slgasygeeserlnvveqqlskelrnlavagpsrsgervkfweswrekmqcgfkglsla 606
 QY 617 GSAAAQSLILGMFPSSDGYTLVEENGALIKGKDLCLLTASAMRP 661
 DB 607 gnaatgatlllgmfrpsdgytlvdngltiklgwkqlsiltasawrp 651

RESULT 3
 AAB28569
 ID AAB28569 standard; Protein: 653 AA.
 XX
 AC AAB28569;
 XX

DT 09-FEB-2001 (first entry)
 XX Arabidopsis SCARECROW protein.
 DE
 XX SCARECROW gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
 KW transgenic plant; cell division; molecular marker; herbicide resistance;
 KW salt resistance; pathogen resistance; insect resistance.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W0200053723-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 07-MAR-2000; 2000WO-US05875.
 XX
 PR 10-MAR-1999; 99US-0265585.
 XX
 PA (UYNT) UNIV NEW YORK STATE.
 XX
 PA Benfey PN, Di Laurenzio L, Wyszocka-Diller J, Malamy JE, Pysch L;
 PI Helariutta Y, Bruce W, Lim J;
 DR WPI: 2000-594315/56.
 XX
 DR N-PSDB; AAC65280.
 XX
 PT Scarecrow gene useful for producing transgenic plants expressing genes
 PT whose product increases starch, lignin or cellulose biosynthesis and
 PT confers herbicide, pathogen or insect resistance -
 XX
 PS Claim 2; Fig 5A; 200pp; English.

The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MOTIF III (WHIRL) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryos and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.

Sequence 653 AA;

Query Match 53.1%; Score 1861; DB 21; Length 653;
 Best Local Similarity 57.3%; Pred. No. 6.2e-139;
 Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

QY 4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHLHLDHQAAMAAWVKRP 56
 DB 12 PPPHSPILRTSSGSSSSNNRPPP--PPPP-----LVMVTKRL 48
 QY 57 ASDM-----DLPPRR--H-VTGDLSVT-----AAAAGVGGSGAPS---SASAQ 96
 DB 49 asemssnpdymnsrrprvshllidsnyntvtpqgppltaaatvssqpnplsvcgfsg 108
 QY 97 LPLPTQLHQLPRAFQNHAPVNDVPAHPAPAAHQAAGEATASTTAWDGIIRDIIGSSG 156
 DB 109 lpfvfrs-----drggrtnvmsvqpmnd---qdssssasptvwdaifrdllhss- 154
 QY 157 GAAVSITQLIHNVRELIHPCNPGLASILELRSLAADPA-----PLP----- 200
 DB 155 -tsvisipqlignvrdlflfcpnplgallelyrltsimldpsssdsbpqtfelyqismn 213
 QY 201 --PPPOQHALLHGAPAAAPAGTLPPPPPLDPRKRHEHPPQOOQOOEHPHAPQSK 258

```

Db 214 pspqqqqqhq-----:qgqqqkpppppqqgqere---nsstd 248
QY 259 APTAETAAAA-AAQAAAAAAKKEKQRRKQDEBGLHLLTLLOCAEVAMDND 317
Db 249 appqpetvatvavqtnlaaalrerkkeikrqkqdeqihlllllqcaevsadnl 308
QY 318 AHQTLLEIAELATPECTSTQRYAAVFAEAMARLVSSCIGLAPPPGSPARARHG-RV 376
Db 309 ankllleisqlstpytsaqryaayfseamsarllinsciglyaalp--srmppqchslkm 366
QY 377 AAAFVFNCGISPPVKFSHTANQAOEAFEREERVHIIDLDIMQGLQWPGFPHILASRP 436
Db 367 vsafgyfngisplytkfshltanqalqeaekedsvhlldldimglqwpqglfhlilasp 426
QY 437 GPPRVLTGLGASMEALATGKRLSPADTLGPREFCVAKANVDPEKGVRRRAV 496
Db 427 gpprvltlgltsmealqatgkrlsdlfclgiprfcplakvgnldterlnvkrav 486
QY 497 AVHMLHSLYDVYTGSDSNLTAMLIQRLAPKVTMVEDLSHSGFLARFEVAIHYSALFD 556
Db 487 avhmlqhslydvrtgsdahnltwlqlrapkvvtveqgdishagsflgrfveahysalfd 546
QY 557 SLDASTGSDSPERHVVQQLSRETRNVILAVGPARTDVKGFGVREKLAQSGFPAASIA 616
Db 547 slgasygeeserhveqqlskeltnvllavgsprsgvktreswrekmqcgfkqisla 606
QY 617 GSAAGASLLGMPSPDGYLVEENGALKGWDCILLTASWMP 661
Db 607 gnaatqatlilgmfpdgytlvdhngclkgwkdislitasawcp 651

```

RESULT 4

ID AAM38192 standard: Protein; 716 AA.

AC AAM38192;

DT 11-MAY-1998 (first entry)

XX Maize ZCARECROW ZCR protein.

XX ZCARECROW; ZCR; SCARECROW; SCR; transgenic plant; root;

KW gravitropism; crop improvement; maize.

OS Zea mays.

XX Key Location/Qualifiers

```

FT 7 Misc-difference 7 /note= "encoded by TAA"
FT 9 Misc-difference 9 /note= "encoded by TAA"
FT 15 Misc-difference 15 /note= "encoded by TGA"
FT 19 Misc-difference 19 /note= "encoded by TGA"
FT 20 Misc-difference 20 /note= "encoded by TGA"
FT 26 Misc-difference 26 /note= "encoded by TGA"
FT 31 Misc-difference 31 /note= "encoded by TGA"
FT 34 Misc-difference 34 /note= "encoded by TGA"
FT 35 Misc-difference 35 /note= "encoded by TGA"
FT 38 Misc-difference 38 /note= "encoded by TGA"
FT 46 Misc-difference 46 /note= "encoded by TGA"
FT 48 Misc-difference 48 /note= "encoded by TGA"
FT 62 Misc-difference 62 /note= "encoded by TGA"

```

```

FT 65 /note= "encoded by TGA"
FT 68 /note= "encoded by TAG"
FT 82 /note= "encoded by TAG"
FT 113 /note= "encoded by TAG"
FT 114 /note= "encoded by TGA"
FT 134 /note= "encoded by TAG"
FT 136 /note= "encoded by TGA"
FT 148 /note= "encoded by TGA"
FT 184 /note= "encoded by TAA"
FT 188 /note= "encoded by TAA"
FT 190 /note= "encoded by TAG"
FT 239 /note= "encoded by TAG"
FT 245 /note= "encoded by TAG"
FT 259 /note= "encoded by TGA"
FT 316 /note= "encoded by TAA"
FT 318 /note= "encoded by TGA"
FT 378 /note= "encoded by TGA"
FT 386 /note= "encoded by TAG"
FT 586 /note= "encoded by TAG"
FT 615 /note= "encoded by TAA"
FT 625 /note= "encoded by TGA"
FT 669 /note= "encoded by TGA"
FT 671 /note= "encoded by TAG"
FT 674 /note= "encoded by TAA"
FT 694 /note= "encoded by TGA"
FT 697 /note= "encoded by TAG"
FT 707 /note= "encoded by TAG"
FT 318..539 /note= "encoded by TGA"
FT /note= "region of homology to SCR"

WO9741152-A1.
PD 06-NOV-1997.
PF 25-APR-1997; 97WO-US07022.
PR 24-APR-1997; 97US-0842445.
PR 26-APR-1996; 96US-0638617.
XX (UYNV ) UNIV NEW YORK STATE.
XX Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
XX Pysh L, Wysocka-Diller J;
DR WPI: 1997-549683/50.
XX N-PSDB; AAT95767.
PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
PT cell division and therefore alter root development, or alter plant

```


PT stem or hypocotyl gravitropism
XX
PS Claim 2; Page 138-140; 221pp; English.
XX
CC This is the deduced amino acid sequence of the maize ZCARECROW gene
CC (see AAT95767). It has a region of homology to Arabidopsis SCARECROW
CC (SCR) protein (see AAM38178). SCR is a member of a novel protein
CC family and contains a number of potential functional domains
CC similar to those found in transcription factors. SCR is expressed
CC specifically in embryo root progenitor tissue and in certain root
CC and stem tissues. It regulates a specific asymmetric division, and
CC controls gravitropic response in aerial structures and root
CC formation. SCR proteins (see AAM38178-201) from dicots and monocots,
CC SCR proteins lacking 1-4 of MOTIFs I-VI, and polypeptides
CC corresponding to MOTIF I, II, III, IV, V or VI of SCR are claimed.
CC Transgenic plants can be engineered to overexpress the SCR protein,
CC so that cell division is increased in roots, resulting in thicker
CC root development, while a plant with an altered stem or hypocotyl
CC gravitropism is less susceptible to lodging. Plants that contain
CC an antisense molecule that suppresses the expression of endogenous
CC SCR gene product have thinner root development. A gene of interest
CC can be placed under control of a SCR promoter and expressed in a
CC plant to confer herbicide, salt, pathogen or insect resistance, or
CC when expressed in stems to increase starch, lignin or cellulose
CC biosynthesis (all claimed).
XX
SQ Sequence 716 AA:

Query Match 32.3%; Score 1131.5; DB 18; Length 716;
Best Local Similarity 78.1%; Pred. No. 4.2e-81;
Matches 225; Conservative 12; Mismatches 24; Indels 27; Gaps 3;

QY 234 RHEHPPPCQQQQQDEPHAPQSPKAPTAETAAAAAQAQAAAAAAKERKEORRQRD 293
Db 279 rrrsrp-----rrpprrpppprr--sgrrcsgssatrrastcx----- 316
QY 294 EEGHLLTLTLLOCAEAVNANDDAHOTLLEIAELATPFGSTORVAAYFAEAMSARLV 353
Db 317 -----vllllqcaevnaddahqtllelaelaprgtscqvaayfaeamsarivs 371
QY 354 SCLGLYAPLPSPGSPAARLRHGVAAAFQVFNQISPFVKFSHTANAOIQAEFEREVRHI 413
Db 372 sclgilyaplpsspsaarlrhgrvaafqvfngispfvkfshtanqalqeateerevhl 431
QY 414 IDLDIMQGLQWPGLEFHILASRPGSPRRVRLTGLASMEALEATGKRISDPADTLGLPEF 473
Db 432 idldimglqwpglfhlilvsrpgsprvrltglasmdaleatgkrlsdftadtlglpfe 491
QY 474 CAAVEKAGNVDEPKLGTREAVAVHMLHSLDYDGSNSTIMLQIR 521
Db 492 cavaekagnvdpkrlgvtreaavahwphnslidydgssdntlwlilqr 539

RESULT 5
AAB28591
ID AAB28591 standard; Protein; 716 AA.
AC AAB28591;
XX
DT 09-FEB-2001 (first entry)
XX
DE Maize ZCR protein.
XX
KM Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
KM transgenic plant; cell division; molecular marker; herbicide resistance;
KM salt resistance; pathogen resistance; insect resistance.
XX
OS Zea mays.
XX
PN MO200053723-A2.
XX
PD 14-SEP-2000.

XX
XX 07-MAR-2000; 2000WO-US05875.
XX
XX 10-MAR-1999; 99US-0265585.
XX
PR (UWNY) UNIV NEW YORK STATE.
XX
PA
PI Benfey PN, Di Laurenzio L, Wyszoka-Diller J, Malamy JE, Pysh L;
PI Helaridutta Y, Bruce W, Lim J;
DR WPI; 2000-594315/56.
DR N-PSDB; AAC65298.
XX
PT Scarecrow gene useful for producing transgenic plants expressing genes
PT whose product increases starch, lignin or cellulose biosynthesis and
PT confers herbicide, pathogen or insect resistance
XX
PS Claim 2; Fig 17B; 200pp; English.
XX
CC The present sequence is given in a specification relating to the
CC structure and function of a regulatory gene, SCARECROW (SCR).
CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
CC sequence similar to the sequence of MOTIF III (VHIII) of Arabidopsis SCR
CC protein are disclosed. SCR and SCL genes are useful for producing
CC transgenic plants whose cell division is modified and root and/or stem
CC development and gravitropism of stem or hypocotyl is altered. Cell
CC division is increased or decreased in roots resulting in thicker or
CC thinner root development. The transgenic plants are useful for expressing
CC a gene of interest encoding a gene product that confers herbicide, salt,
CC pathogen or insect resistance in root or embryo and genes encoding
CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
CC confers less susceptibility to lodging in the transgenic plants than a
CC wild-type plant. SCR gene sequences are also useful as molecular markers
CC for a quantitative trait e.g. root or gravitropism trait in molecular
CC breeding of crop plants.
XX
SQ Sequence 716 AA:

Query Match 32.3%; Score 1131.5; DB 21; Length 716;
Best Local Similarity 78.1%; Pred. No. 4.2e-81;
Matches 225; Conservative 12; Mismatches 24; Indels 27; Gaps 3;

QY 234 RHEHPPPCQQQQQDEPHAPQSPKAPTAETAAAAAQAQAAAAAAKERKEORRQRD 293
Db 279 rrrsrp-----rrpprrpppprr--sgrrcsgssatrrastcx----- 316
QY 294 EEGHLLTLTLLOCAEAVNANDDAHOTLLEIAELATPFGSTORVAAYFAEAMSARLV 353
Db 317 -----vllllqcaevnaddahqtllelaelaprgtscqvaayfaeamsarivs 371
QY 354 SCLGLYAPLPSPGSPAARLRHGVAAAFQVFNQISPFVKFSHTANAOIQAEFEREVRHI 413
Db 372 sclgilyaplpsspsaarlrhgrvaafqvfngispfvkfshtanqalqeateerevhl 431
QY 414 IDLDIMQGLQWPGLEFHILASRPGSPRRVRLTGLASMEALEATGKRISDPADTLGLPEF 473
Db 432 idldimglqwpglfhlilvsrpgsprvrltglasmdaleatgkrlsdftadtlglpfe 491
QY 474 CAAVEKAGNVDEPKLGTREAVAVHMLHSLDYDGSNSTIMLQIR 521
Db 492 cavaekagnvdpkrlgvtreaavahwphnslidydgssdntlwlilqr 539

RESULT 6
AAG48516
ID AAG48516 standard; Protein; 405 AA.
AC AAG48516;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61274.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-013629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147004;
PR	04-AUG-1999;	9905-0147208;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147260;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147935;
PR	10-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148319;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148655;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	17-AUG-1999;	9905-0149175;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149723;
PR	20-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;
PR	25-AUG-1999;	9905-0150566;
PR	25-AUG-1999;	9905-0150084;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151303;
PR	31-AUG-1999;	9905-0151338;
PR	01-SEP-1999;	9905-0151930;
PR	01-SEP-1999;	9905-0152633;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153758;
PR	13-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154079;
PR	20-SEP-1999;	9905-0154739;
PR	22-SEP-1999;	9905-0155133;
PR	23-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-0156658;
PR	29-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157753;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158232;
PR	12-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0158983;
PR	13-OCT-1999;	9905-0159294;
PR	13-OCT-1999;	9905-0159295;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160041;
PR	21-OCT-1999;	9905-0160041;
PR	21-OCT-1999;	9905-0160076;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160777;
PR	22-OCT-1999;	9905-0160981;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;

PR 26-OCT-1999; 99US-0161361
PR 28-OCT-1999; 99US-0161920
PR 28-OCT-1999; 99US-0161992
PR 28-OCT-1999; 99US-0161993
PR 29-OCT-1999; 99US-0162142

Query Match	20.1%;	Score 705.5;	DB 21;	Length 246;
Best Local Similarity	57.6%;	Pred. No. 5.5e-48;		
Matches 141;	Conservative 37;	Mismatches 60;	Indels 7;	Gaps 3;

QY 419 MÖGLÖWPGLPHILASRPGGPPRVRLTGLGASMEALEATGKRLSDPADTLGLPFECAVAE 478

Db 1 mgqlqpalfhllasrprklrsirtgygssdllastryladfasslnlpfehnpieg 60

QY 479 KAGN-VDPEKIGVTRREAAVAHMLHLSLDYTGSDSNTLWLIQRLAPKVVTMEQDLSH- 536

Db 11gn11qpsqlatrggeaavvnhmmgnrlydvfgnmetlellrllkpnlltvegeatsya 120

[illegible]

506 WZROGNIENIE KOSCIOLU I CIERPIECIOWNI KSIĘKWIŁCIŁT 655

178 -krmkw[ee]srvafrvns]rnnatag[]am]hwnavt]veenat]r]awkd[]s[]t 236

OV 656 ASAWR 660

Db 237 asawk 241

RESOLVED
AYY02540

[illegible]

XX	16-JUL-1999	(first entry)
DT		

Protein encoded by wheat Rht clone 5a1 genomic sequence

AA Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;
KW

KW
paclobutrazol.

OS Triticum aestivum.
xy

PN W09909174-A1.
XX

PD 25-FEB-1999.
XX

PF 07-AUG-1998; 98WO-GBU2383
XX

PK 13-AUG-1991; 9/GB-001/192
XX

XX
XX
XX

XX
DB WBT: 1999-191040/15

DR N-PSDB; AAX36279.
XX

PT New *Triticum aestivum* polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is

XX PI antagonised by gibberellin, used to confer a dwarf phenotype

XX
XX
XX

CC its homologues) that, when expressed in *Triticum Aestivum*, inhibit growth of the plant, this growth inhibition is antagonised by CC

CC plants, conferring a dwarf phenotype on a plant which is correctable

by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by the wheat Rht clone 5a1 genomic sequence.

Sequence 623 AA:

Query Match 19.1%; Score 670; DB 20; Length 623;
Best Local Similarity 30.9%; Pred. No. 1,2e-44;
Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

```

QY 79 AAAGVGGSGAPSSASAOPLALPTQLHQLPRAFGHNAPEVDVPAHPARAHAQAGGATA 138
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 aqsgsg99gmgsedkmmvs-----aaagegeevd-----ellaalgykxra 50

QY 139 STTAWVDGIIRDI-----IGSSG---GAAYSIQILHNVEIHPNCPLASLLELRSL 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 51 scmadvaqkleglemamgmgygagaapddsfathlatdtyhprtdlssvvesml-sel 109

QY 193 AADPAFLPPPO-----PQHALLHGAPAAAPAGLTP---PP 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 nappplppaplnaststvgsgyfdlppsvdsssiyalpripapatacpadisa 169

QY 228 PPLPRKRHNPPPCQQQQQDEPPHAPDSPAFAETFAAANAQAANAARERKEEO 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 170 dsavipkmttvgssstssssssslgggavveaaprydaanaapralprrvv----- 224

QY 288 RRKQDEGLMLLTLLQCAEAVNADNDDAHQTLETAELATPEFGSTOVAAVFAEAM 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 225 --vdtgeagirlthallacaeavqgenisaeealykqilpilaasqggamtkvaaylgeal 282

QY 348 SARLVSSCLGYAPLPDGS---PAAARLHGRVAAAFQVNGISPEVFKSHFTANQAOIE 403
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 283 arvrfr-----frppdslldaafadllhab-----fyescpylkfhftangalle 330

QY 404 AFERERVIHLDLMQGLWPGFLHILASRPGPRRRLTGLCA---SKEALEATGKR 459
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 atagcrryvvvdfgikgwpallqalalrpgppsfirllgypppddetdalqyvwk 390

QY 460 LSDFADTLGLPEFCFCAV-----EKAGNVDPKRGVRRREAVAHMHL--HSL 505
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 391 laqfahltlrvtqyrglvaatladeptfmigpegeedpne-----epeviavnsyfemhrl 446

QY 506 YDVTGSDSNTLMILIORLAPKVVTWVEDLSH-SGSFLARFVEALHYYSALFDSLDAASYGE 564
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 447 laqpalalekltgtvraavprlvtvgeaahnsqtlfdrtfeshlyysltmdstlegssg 506

QY 565 DSPER-----HVEQQLSREIRNVLAAGFAAR-GDVKRGSMREKLAQS 608
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 507 ggpsevsqgaapaagatdgvmevyylgdlcnvacegeerterhelqgwtnrignaa 566

QY 609 GFRAASLAGSAAQAASLIGMFP-SDGYTLVEENGALKLWKDCLLTASAMR 660
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 gfevthrgsnaykqastllalfaggdgykveekegcltlghvhtprlatsawr 619

RESULT 10
AA02541
ID AA02541 standard; Protein; 630 AA.
XX AA02541:
XX AC
XX AA02541:
XX DT 16-JUL-1999 (first entry)
XX DE Protein encoded by maize 1a1 genomic clone sequence.
XX KW Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
```

KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
XX paclobutrazol; maize.
OS Zea mays.
XX MO9909174-A1.
PN 25-FEB-1999.
XX 07-AUG-1998; 98MO-GB02383.
XX 13-AUG-1997; 97GB-0017192.
XX (PLAN-) PLANT BIOSCIENCE LTD.
PA Harberd NP, Peng J, Richards DE;
PI WPT; 1999-181040/15.
DR N-PDSB; AAX36280.
XX

PT New Triticum Aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 9b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum Aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence is encoded by
CC the maize 1a1 genomic clone sequence.

Sequence 630 AA:

Query Match 19.0%; Score 666; DB 20; Length 630;
Best Local Similarity 31.9%; Pred. No. 2.6e-44;
Matches 209; Conservative 80; Mismatches 253; Indels 114; Gaps 21;

```

QY 71 GDLSDVTAANAAGVGGSGAPSSASAOPLALPTQLHQLPRAFGHNAPEVDVPAHPARAHA 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16 gskdkmmaaaagag-----eqeedvd-----ella 42

QY 131 QACGEATASTTAMVDGIIRDI-----IGSSGGAANSITQ--LIHNVEIHPNCPLASL 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 algykvrssdmadvakleqlemamgmgygagagadqfvshtlatdvhypsdllsw 102

QY 184 LELRLSLLAADPAFLPP-PPQQHALLHGAPAAAPAGLTPRRPRLDKRRNH----- 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 vesml-selnapparpripptprlaststvtsgaaagayfdlppravdssssyalkp 161

QY 239 -PPCCQQQQQDEPPHAPDSPAFAETFAA-----AAAAA 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 lpsvpapad--peldsarepkrmttvgstssssssssmdgtrftrsvveaapat 218

QY 273 QAAAAAARERKEEQRKRDDEGLMLLTLLQCAEAVNADNDDAHQTLETAELATPFP 332
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 219 qasaaanpavrvvvvdtq--eaqrlvhalacaeavqgenfsaaealvxqjplmlassg 276

QY 333 GTSTORVAAYFAAMSARLVSSCLGYAPLPDGSFAAARLHGRVAAAFQV-FNGISPFVK 391
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 ggamtkvaaylgealarr-----vytrfrp--pdslldaafadllhanfyescpylk 327

QY 392 FSHFTANQAOIEAFERERVIHLDLMQGLWPGFLHILASRPGPRRRLTGLCA--- 448
```

```
Db      328 fahfntaagallleaafagerrrhvvdffikgmmqpalldgaalrpggppsfritlqvppqp 387
QY      449 -SMEALFATGKRSLSDPADLTGLDFEECA-VAERAGVNDPEKIGV-----TRREAVAVHM 500
Db      388 decdaqlqvgwklagfahftrvdfgyrlvaatladalepmlqpegddcddepevalavns 447
QY      501 LH--HSIYDVTGSDSNLMLIQRLAKRYVMVQDLSH-SGSLAFVFAIHYSLAFDS 557
Db      448 vfelhrliaagpalekvylgtvraivrprilyvegeanhsgtlfidfteslhytslmdfs 507
QY      558 L-----DASYGEDSPERHVVEQQLSPREIRNVLAVGSPAPR-GDVKFGSMREKI 605
Db      508 legagagsgstaspaagaagtdqvmsevylygticvvaacegaetehetlqgwrsrl 567
QY      606 AOSGFRAASLAGSAAQAQSLILGMFP-SDGYTLVEENGALIKLGMKDLCLLTASAMR 660
Db      568 ggsqgfapvhlgsnaykqasllalfagsgdgyrveekdgcrltqwhrplatsawr 623
```

RESULT 11

AAG22966 standard; Protein; 240 AA.

AAG22966;

XX AC AAG22966;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 26096.

XX KM Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139765.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR	02- AUG -1999	99US-0146538
PR	03- AUG -1999	99US-0147038
PR	04- AUG -1999	99US-0147204
PR	04- AUG -1999	99US-0147192
PR	05- AUG -1999	99US-0147302
PR	05- AUG -1999	99US-0147260
PR	05- AUG -1999	99US-0147736
PR	06- AUG -1999	99US-0147416
PR	06- AUG -1999	99US-0147416
PR	09- AUG -1999	99US-0147493
PR	09- AUG -1999	99US-0147935
PR	10- AUG -1999	99US-0148171
PR	11- AUG -1999	99US-0148341
PR	12- AUG -1999	99US-0148341
PR	13- AUG -1999	99US-0148365
PR	13- AUG -1999	99US-0148364
PR	16- AUG -1999	99US-0149376
PR	17- AUG -1999	99US-0149175
PR	18- AUG -1999	99US-0149426
PR	20- AUG -1999	99US-0149722
PR	20- AUG -1999	99US-0149723
PR	22- AUG -1999	99US-0149902
PR	23- AUG -1999	99US-0149930
PR	23- AUG -1999	99US-0150566
PR	26- AUG -1999	99US-0150884
PR	27- AUG -1999	99US-0151065
PR	27- AUG -1999	99US-0151065
PR	27- AUG -1999	99US-0151066
PR	30- AUG -1999	99US-0151303
PR	31- AUG -1999	99US-0151338
PR	01- SEP -1999	99US-0151330
PR	07- SEP -1999	99US-0152363
PR	10- SEP -1999	99US-0153070
PR	13- SEP -1999	99US-0153758
PR	15- SEP -1999	99US-0154018
PR	16- SEP -1999	99US-0154039
PR	20- SEP -1999	99US-0154719
PR	22- SEP -1999	99US-0155133
PR	23- SEP -1999	99US-0155586
PR	24- SEP -1999	99US-0155549
PR	28- SEP -1999	99US-0156458
PR	29- SEP -1999	99US-0156596
PR	04- OCT -1999	99US-0157153
PR	05- OCT -1999	99US-0157757
PR	06- OCT -1999	99US-0157865
PR	07- OCT -1999	99US-0158032
PR	08- OCT -1999	99US-0158232
PR	12- OCT -1999	99US-0158369
PR	13- OCT -1999	99US-0158933
PR	13- OCT -1999	99US-0159294
PR	13- OCT -1999	99US-0159292
PR	14- OCT -1999	99US-0159329
PR	14- OCT -1999	99US-0159330
PR	14- OCT -1999	99US-0159331
PR	14- OCT -1999	99US-0159637
PR	14- OCT -1999	99US-0159638
PR	14- OCT -1999	99US-0159684
PR	18- OCT -1999	99US-0160741
PR	21- OCT -1999	99US-0160767
PR	21- OCT -1999	99US-0160768
PR	21- OCT -1999	99US-0160770
PR	21- OCT -1999	99US-0160814
PR	21- OCT -1999	99US-0160815
PR	22- OCT -1999	99US-0160880
PR	22- OCT -1999	99US-0160981
PR	22- OCT -1999	99US-0160981
PR	25- OCT -1999	99US-0161404
PR	25- OCT -1999	99US-0161405
PR	26- OCT -1999	99US-0161406
PR	26- OCT -1999	99US-0161356
PR	26- OCT -1999	99US-0161360
PR	26- OCT -1999	99US-0161361
PR	28- OCT -1999	99US-0161920

PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	18.6%;	Score 651;	DB 21;	Length 240;
Best Local Similarity	56.5%;	Pred. No. 1.1e-43;		
Matches 134;	Conservative 37;	Mismatches 58;	Indels 8;	Gaps 4

```
OY 427 LEFLIASHRGPPRRRLTIGASMEALEANGKRLSDPADNLTGJPFEECAAEAKAGN-VDP 465S
Db 4 lfhliastrprkrlstirrlgissdllastryrrladiasslnlpfethrlegllaylda 63
OY 486 EKLGVTRREAAVAHMLHNLSLVDVGDGSNTLMTLQRLAPKVVWVEEDLSH--SGSPLAR 543S
Db 64 sqldtgrgeaxvnmtenrlrydvngndetelrtlrknlitvveqelnyddgsflgr 123S
OY 544 FVEALHYHSLEFDSLIDASYGEDSPERHAYVEEQLLSREIRNVLAYGGPARTGDVKFGSGRE 603S
Db 124 fvealhyssalfdalgydlgeesgerfveeglvlt-elmvlahgggr----krmkwke 178S
OY 604 KLAOSGFPAASLAGSAANAQASLLGLMRPDSGYTYLVEENGALKLGWKDLCLITISAMR 660S
Db 179 elstvgfrpvslgnpatqaelglgmipmgylgtveengtlrgywkdsjlltsasawk 235S
```

RESULT 12

ID AAY02539 standard; Protein; 425 AA.

AC AAY02539;

DT 16-JUL-1999 (first entry)

DE Protein encoded by the wheat Rht clone C15-1 cDNA sequence.

KW Rht gene; homologue: *Triticum aestivum*; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol.

OS *Triticum aestivum*.

PN W09909174-A1

PD 25-FEB-1999

PF 07-AUG-1998; 98WO-GB02383

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

XX

PT provides inhibition of the growth of plants, which inhibition is

XX

PS Disclosure; Fig 7b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologue gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol


```
QY 438 PPRVRLTGLA----SMEALEATGKRLSDPADTLGLPFEFCA-VAEKAGNVDEKLGV-- 490
|| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 pPfrlItlgIppapdnshlhevqcklaqlaealvefeyrgfvansladidsmlleirp 409
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 TRREAVAVHMLH--HSLYDVTGSDSNTLMLIORLAPKVYVMEBODLSHSGS-FLARFVEA 547
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 sdeavavnsvfelhklIgrpgiekvlgvqkIkpvlftvegesnngpvlfdlrfes 469
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 548 IHYSALFSDSLDASYGEDSEPRHVVVEQQLSREIRNVLAVGSPART-GDVKFGSREKLA 606
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 lhystylfidsle---gvpsnqdkvmsevyIqkqIcnlvacegpdvverhetlsqgnrlf 526
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 607 QSGFRAASLAGSAAQAQSLILGMFPS-DGYTLVEENGALTKGMDCLLTASAMR 660
|| || || || || || || || || || || || || || || || || || || || ||
Db 527 ssgIapahlgnsaIfkqasmllsvfnsggyrveesngclmIghwtrpIlttsawk 581

RESULT 14
AAE01892
ID AAE01892 standard; Protein; 587 AA.

AC AAE01892;
DT 31-JUL-2001 (first entry)
DE Arabidopsis thaliana transcription factor, G307.
KW Transcription factor; biochemical characteristic; controlling element;
KW structural characteristic; developmental characteristic; gene therapy;
KW agricultural biotechnology; plant trait modification.
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Domain 323..339
FT /label= Conserved_domain

XX
PN WO200136597-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31344.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (CREE/) CREELMAN R.
PA (YUGG/) YU G.
PA (ADAM/) ADAM L.
PA (RIEC/) RIECHMANN J L.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMARA R.
PA (PILG/) PILGRIM M.
PA (PINE/) PINEDA O.
PA (JIANG/) JIANG C.
XX
PI Creelman R, Yu G, Adam L, Riechmann JL, Heard J, Samara R;
PI Pilgrim M, Pineda O, Jiang C;
XX
DR N-PSDB; AAD05776.
XX
WT: 2001-335999/35.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the biochemical characteristics of plants e.g. corn,
PT potato and cotton plants -
XX
PS Claim 4; Page 68-70; 127pp; English.
XX
CC The present sequence is Arabidopsis thaliana transcription factor,
CC G307. The transcription factor is used for altering a plant's
```

```
CC biochemical characteristics. The transcription factor may be used to
CC alter the structure and developmental characteristics of plants such as
CC soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,
CC alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,
CC raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,
CC grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,
CC pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon,
CC rosaceae fruits and/or vegetable brassicas. Transcription factors are
CC key controlling elements of biological pathways and altering expression
CC levels of 1 or more transcription factors can change entire biological
CC pathways in an organism. Therefore manipulating transcription factor
CC levels in plants offers great potential in agricultural biotechnology
CC for modifying a plant's traits. Transcription factor cDNA is useful in
CC gene therapy.
XX
SQ Sequence 587 AA;

Query Match 17.0%, Score 594.5; DB 22; Length 587;
Best Local Similarity 31.8%; Pred. No. 1,le-38;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

QY 165 LIHNVREITHPCNG-LASLELRSLAADPAPLPPOPOQHALLHGAPAAPAGL- 222
|| || : || || || || || || || || || || || || || || || || || || || || ||
Db 81 lshlatlvh-ypselyswldmll-----selpprlp-----assngld 120

QY 223 -TLPPP-----PRLPKRRHNEHPCCQOQOQEEPRAPQSKAPTAEETAA- 268
|| || : || || || || || || || || || || || || || || || || || || || || ||
Db 121 pvlrpselgfrasdylkvlrpnalylgfraldssssnnqkrlkscspsdmwstst 180

QY 269 -----AAAAQAAAAAAKKEKEORRKRORBEGLHLLTLLOCAEAVNADNLDDA 318
|| || : || || || || || || || || || || || || || || || || || || || || ||
Db 181 grqigvgigtvtltttttaaestrsylvlvdsgengvrlvhalmaeaeilqgnllcia 240

QY 319 HQTLEIAELATPFGSTNRVAAYFAEAMASRLVSSCLGLVAPLPGPSAAARLHGRVAA 378
|| || : || || || || || || || || || || || || || || || || || || || || ||
Db 241 ealvkgIgcIavsgagamtkvaltyfaealarr-----Iylspgn-----qldhclsd 289

QY 379 AFQV-FNGISPFVKFSHTFANOAIQEAEREBRVHIIIDIMOGIQLWPGFLHILASRPG 437
|| || : || || || || || || || || || || || || || || || || || || || || ||
Db 290 tIqmhlyetcpylkfhfhfanganqllaefegkkrvhlidtsmngqIqwpalmagallregg 349

QY 438 PPRVRLTGLA----SMEALEATGKRLSDPADTLGLPFEFCA-VAEKAGNVDEKLGV-- 490
|| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 pPfrlItlgIppapdnshlhevqcklaqlaealvefeyrgfvansladidsmlleirp 409
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 TRREAVAVHMLH--HSLYDVTGSDSNTLMLIORLAPKVYVMEBODLSHSGS-FLARFVEA 547
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 sdeavavnsvfelhklIgrpgiekvlgvqkIkpvlftvegesnngpvlfdlrfes 469
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 548 IHYSALFSDSLDASYGEDSEPRHVVVEQQLSREIRNVLAVGSPART-GDVKFGSREKLA 606
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 lhystylfidsle---gvpsnqdkvmsevyIqkqIcnlvacegpdvverhetlsqgnrlf 526
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 607 QSGFRAASLAGSAAQAQSLILGMFPS-DGYTLVEENGALTKGMDCLLTASAMR 660
|| || || || || || || || || || || || || || || || || || || || ||
Db 527 ssgIapahlgnsaIfkqasmllsvfnsggyrveesngclmIghwtrpIlttsawk 581

RESULT 15
AAG38577
ID AAG38577 standard; Protein; 517 AA.
AC AAG38577;
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47612.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
```

OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153756.

[illegible]

```

RESULT 2
US-09-117-853-8
; Sequence 8, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jinrong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-8

```

[illegible]

```

? RESULT 3
? US-09-117-853-6
? Sequence 6, Application US/09117853
? Patent No. 6507126
? GENERAL INFORMATION:
? APPLICANT: Harberd, Nicholas P
? APPLICANT: Peng, Jinrong
? APPLICANT: Carol, Pierre
? APPLICANT: Richards, Donald E
? TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
? FILE REFERENCE: 620-45
? CURRENT APPLICATION NUMBER: US/09/117, 853
? CURRENT FILING DATE: 1998-08-12
? EARLIER APPLICATION NUMBER: PCT/GB97/00390
? EARLIER FILING DATE: 1997-02-12
? EARLIER APPLICATION NUMBER: GB 9602796.6
? EARLIER FILING DATE: 1996-02-12
? NUMBER OF SEQ ID NOS: 12
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 6
? LENGTH: 259

```

```

; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-6

```

Query Match	5.68;	Score 196;	DB 4;	Length 259;
Best Local Similarity	38.28;	Pred. No. 2.7e-08;		
Matches 47;	Conservative 24;	Mismatches 40;	Indels 12;	Gaps 2;

[illegible]

```

1      RESULT      4
2      US-09-041-886-23
3      ; Sequence 23, Application US/09041886
4      ; Patent No. 6235872
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Bredesen, Dale E.
7      ; APPLICANT: Rabilzadeh, Sharoz
8      ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
9      ; TITLE OF INVENTION: Polypeptides and Methods of Use
10     ; NUMBER OF SEQUENCES: 72
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Campbell & Flores LLP
13     ; STREET: 4370 La Jolla Village Drive, Suite 700
14     ; CITY: San Diego
15     ; STATE: California
16     ; COUNTRY: United States
17     ; ZIP: 92122
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/09/041,886
25     ; FILING DATE:
26     ; CLASSIFICATION:
27     ; ATTORNEY/AGENT INFORMATION:
28     ; NAME: Campbell, Cathryn A.
29     ; REGISTRATION NUMBER: 31,815
30     ; REFERENCE/DOCKET NUMBER: P-LJ 2626
31     ; TELECOMMUNICATION INFORMATION:
32     ; TELEPHONE: (619) 535-9001
33     ; TELEFAX: (619) 535-8949
34     ; INFORMATION FOR SEQ ID NO: 23:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 1185 amino acids
37     ; TYPE: amino acid
38     ; TOPOLOGY: linear
39     ; MOLECULE TYPE: protein
40     ; US-09-041-886-23

```

```

Query Match Similarity 5.1%; Score 179.5; DB 4; Length 1185;
Best Local Similarity 21.5%; Pred No. 4.8e-06;
Matches 166; Conservative 69; Mismatches 259; Indels 279; Gaps 37

QY 2 PPPPP-----PPLTPYCRCP-----H 21
    |||||      |||
Db 442 PPPPPGGLLANSNAPRFRPSTGAOSTANRSTNNNNHQQOQQOQQOQQOQNN 501
    |||||
QY 22 LPPPPSSFNHFL---LHYLQDLHQEAAAAAMVKRPAQMDLRLPPRRHYT----- 70

```



```

Db 502 GPPPPGAFPPPLEGGSSHHAN--PYAMSPSLGSLRPYGPAAHLPPHPSQVSYSOAGPN 559
QY 71 -----GDSVYTAAMAAAG-----VGGSGAP-----SSASAOPLAPPTOLHQL 107
Db 560 GPPYSSSSSSSSSSSSQSCSHSPSQGOGAPYPPPPYPTVTSATLSTYATVASS 619
QY 108 PPAFOHHAPEVDYVPAHPAANAAGGEATASTTAVYDGIIRDTIGSSGAASVITQLIH 167
Db 620 PAGYKTASP-----GPPPYGKRAPSPAYKATATPPGY-----KPGSPSFRGTGTP 666
QY 168 NVREIHPNPGNLASLELRSLAADPAPLPPPPPOQHALLHGAPAAPGL--TLPP 226
Db 667 GYRGTSPGAGTFFK-----PGSPYVGPGLPP-----AGPSGLPSLPP 705
QY 227 PPLPDRKRHHPPCCQOOOEEPP--HPAPOSPAPT-----261
Db 706 PPAAP-----ASGPLSATQIKQFAEYETPESVPPAPSPSPKVVDPVSHASQSAF 761
QY 262 -----AETAAAAAQAAMAAAKERKEEORRRKORDEE 295
Db 762 NKHLDRGFNSCARSDLYFVPLEGSKLAKKRADLYEKVRRAEQRAEREREREKE 821
QY 296 -----GHLTLTLLOCAEA-VNADNLDA-HQTLLEFIELATPF--GTSTQVAAIFYA 344
Db 822 REREKERELERSVKLAQEGRAVPCPSLGPVPHR-----PPEPPGSAVATVPYLG 872
QY 345 -EAMASRLVSSCLGLYA--PLPPGSPAARLH-----GRVAAAFQVFNGISPPVFESH 394
Db 873 PDFTALKTLSF--YARPHMSPGN--RNHFYVPLGAVDPGLGLYN--VPALXSSD 922
QY 395 FTANAOIAOEAFERER-----VHITDDIMOGIOWPGLFHT--LASRPGGPPRY 441
Db 923 PAAREEREAERDLRLKRGFEVKKPELEPLHGVPGGLDPFRHGLALOPGP--979
QY 442 RLTLGLASMEALATGKRLSDFATLGLPFEFCVAEAKAGV-----DDEKLGVPTRREA 495
Db 980 -----GLHPPFPHPSLG-PLBRERKLAAGALRPDMSVYERLAEKQHA 1023
QY 496 VAVHMLHSLDYVGSNTLMQLQRLAPKYVTVWEODLSHSGSFL--ARVEALHYSA- 553
Db 1024 ERVAGL-----GNP-----LARIOMLVNPHHHOHSHLHLHOODATHAASAS 1070
QY 554 ---LPDSL-DASYGEDSP-----ERHNVEOQLSPRITRVLA 586
Db 1071 VHPRLIDPLASGSHLTRIPYPAGTLPLNPLRPHLEHNEVLRHQLEAFARYDLPA 1123

```

RESULT 5
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext. 513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-945-283-2

Query Match          5.0%; Score 173.5; DB 1; Length 1958;
Best Local Similarity 21.4%; Pred. No. 3e-05;
Matches 100; Conservative 26; Mismatches 124; Indels 217; Gaps 21;

QY 2 PPPPP-----PPLTYCRRCRPPHLP--PPSSPHFLTYLHQ-----40
Db 321 PGPPTSTNNHNNHGGPRTSP--RSTSSHQDPGGGPPSAETH--NNHHODPPGGP 375
QY 41 -----LDHQBAAAAAMVKRPAASMDLPPRRRNVYTGDLSDYTAAMAAAGVGSQSPSSAS 94
Db 376 PSTSTNNHNDP-----PGGPPSPPPR-----PSTSS 403
QY 95 ADRLALPQLHOLPRAFOHNAPEVDVPAHPA-----AHAQAGGATASTTAVYDGIIRD 150
Db 404 -----SSHQGPSTRRPPQRRPPRPPSPKISTRAGSENTYQTL-----447
QY 151 IIGSSGAASVITQLIHNVREIT--HPCNPGLASLELRSLIAA-----DPAPLPPRQ 204
Db 448 -----FHSSENKLFSHMGEQEG--DRGTAGGEGDDDRPPSPPR 488
QY 205 PQOHALLHNAAPRAGLTLPPRLPDKRHHNPPRCQOOOEEHPHAPOSP-----257
Db 489 P-----PPLPRLPPPPPPQ-----PPAGSARRRRRGGGGPPRGRR 530
QY 258 ---KAPTAEATAAMAAAO-----AAAAAAKERKEE 286
Db 531 RGGKRRRAEGTEAADADEEDGEDDEDADEBGRDGEGRGAGGGESESESE 590
QY 287 ORRQ--RDEE-----GHLTLTLLOCAEVNADNLDDAHQTLLEIAE 327
Db 591 SSRAEGAPRSABQGVAVGLVVRDGLHDGPERRAGPAAVAADDDLHR-----643
QY 328 LATPFGTSTQVAAATFAEMASRLVSSCLGLYAPLPGPSPAARLHG 374
Db 644 -----GRVLPVLG-----PPGARGFVGLHG 664

```

RESULT 6
US-08-899-595-3
; Sequence 3, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shun
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-595-3

```

```

Query Match 4.9%; Score 171; DB 3; Length 1315;
Best Local Similarity 20.0%; Pred. No. 2,7e-05;
Matches 117; Conservative 53; Mismatches 176; Indels 238; Gaps 25;

```

```

QY 2 PPPPPPLTPYCRRCPPPLP-----PPSSPNHFLHYLHQLDHEAAAAAMVKR 55
DB 656 PPPPPPLPGVCICSSPPSLGGTAISPPPLS----- 689
QY 56 PASDMDLPPRRHVTGDLSDVTAAGVGGAPSSASAPLP---ALPTQLHQLPPAFQ 112
DB 690 --GDATIPPP--PPLPE-----GVGIPSPSS--LPGGTAIP-----PPPL 724
QY 113 HHAPVDVPAHRAAHQAAGGATATTAWDGIIRDIITSSGGAANSITOLIHNVREI 172
DB 725 PGSARIPPPPLP-----GSAG----- 742
QY 173 IHPGNGLASLLELRSLAADPAFLPPPOQHALLHGAPAAAPAGLTPPPPLPD 232
DB 743 IPPPPPLPG-----EAGMPPPP-----LPGGPGLPPPPPPPG 777
QY 233 KRHEHPPCCOOQOQEEP--HPAPOSKPAP-----TAETA- 266
DB 778 GPGIPPPPGMGPPPPFGVPAAPVLPFGLTPRKLYKREVQLRRPNWSKLVAAEDLSQ 837
QY 267 -----AAAAAQAATAAAK-----EKKEBRKRRQDEBGLHL 299
DB 838 DCFWTKVEDRENNELFAKLTLTFSAQTKTSKAKKDOEGGEKKSVQKKVKE----- 891
QY 300 LTLLOCAEAVNADNL-----DDAHOITLEIAELATPFGTSTQORAAVFAEAMS 348
DB 892 ----LKVIDSKTAQNLSTFLGSFRMPYOEIKNVILLEVNE-AVLRESMIONLIKMPPEBQ 946
QY 349 ARLVSSCGLYAPLPGPS-----AARLHGVAAP-----QVFNISIPVKTSHF 395
DB 947 LKMLSELDEYDILAESQFGVWGTVPLRLRPLNALILFKLQFSEQYEN-----IKPEIV 1001
QY 396 TANQAIOAFEREKRVHIIIDIMOGLOWPGLFHILASRPGPPRRVRLTIGASMELEA 455
DB 1002 SVTAACELRKSESEFSNLETTLL-----VGNYNAGSRNAGAGFNISFL-CXLRDTKS 1055
QY 456 TGRRL-----SDFADTLGLPPEFCVAVAEKAGNVDEKL 488
DB 1056 TDQKNTLLHFLAELCENDYPDVYLKFPDELAHV-EKASRVSANL 1098

```

```

RESULT 7
US-09-080-897-2

```

```

; Sequence 2, Application US/09080897
; Patent No. 598574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Weissh, Piri L.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4341
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-080-897-2

```

```

Query Match 4.8%; Score 167.5; DB 2; Length 1248;
Best Local Similarity 19.6%; Pred. No. 4.9e-05;
Matches 114; Conservative 53; Mismatches 171; Indels 243; Gaps 25;

```

```

QY 2 PPPPPPLTPYCRRCPPPLP-----PPSSPNHFLHYLHQLDHEAAAAAMVKR 55
DB 600 PPPPPPL-----PPPLPGGTAISPPPLS----- 625
QY 56 PASDMDLPPRRHVTGDLSDVTAAGVGGAPSSASAPLP---ALPTQLHQLPPAFQ 112
DB 626 --GDATIPPP--PPLPE-----GVGIPSPSS--LPGGTAIP-----PPPL 660
QY 113 HHAPVDVPAHRAAHQAAGGATATTAWDGIIRDIITSSGGAANSITOLIHNVREI 172
DB 661 PGSARIPPPPLP-----GSAG----- 678
QY 173 IHPGNGLASLLELRSLAADPAFLPPPOQHALLHGAPAAAPAGLTPPPPLPD 232
DB 679 IPPPPPLPG-----EAGMPPPP-----LPGGPGLPPPPPPPG 713
QY 233 KRHEHPPCCOOQOQEEP--HPAPOSKPAP-----TAETA- 267
DB 714 GPGIPPPPGMGPPPPFGVPAAPVLPFGLTPRKLYKREVQLRRPNWSKLVAAEDLSQ 773
QY 268 -----AAAAAQAATAAAK-----EKKEBRKRRQDEBGLHL 302
DB 774 DCFWTKVEDRENNELFAKLTLTFSAQTKTKKDOEGGEKKSVQKKVKE----- 824

```



```

1 RESULT 11
2 US-08-718-661-2
3 Sequence 2, Application US/08718661
4 Patent No. 5876972
5 GENERAL INFORMATION:
6 APPLICANT:
7 TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
8 TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation
9 NUMBER OF SEQUENCES: 15
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/718,661
17 FILING DATE:
18 CLASSIFICATION: 530
19 INFORMATION FOR SEQ ID NO: 2:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 667 amino acids
22 TYPE: amino acid
23 TOPOLOGY: linear
24

```



```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-2

```

```

Query Match      4.4%; Score 153.5; DB 4; Length 1298;
Best Local Similarity 20.4%; Pred. No. 0.0007;
Matches 140; Conservative 69; Mismatches 209; Indels 267; Gaps 30;

```

```

QY 53 KRKPAS--DMDLPPRRHVTGDLSDVTAANAAGVSGSPSSASQALPALTQHLQLPRA 110
DB 6 KORGSGPTDGPPTPPSPDRD-----ERGALGKMETEGGDDP-----45
QY 111 FOHNAPEVDVHPAPAAHQAQGEATSTTAWVDGIIRDIISSGGAVSITQLIHNV 170
DB 46 --DHDPD-----HPHDDDARRDGRAPAACTD-----AGEDAGDVSPPQLAL---86
QY 171 EIIHPCNPGIASLLELRSLAADPAFLP-----PPQ 204
DB 87 -----LASVVEEAVRTIPPTDPAASPPRTPAFRADDGDEYDDAADAAGRAPAR 137
QY 205 -POOHALLHGAPAAAGLTPPPPLPDKRRHEHPPRCQQQQQEEBPHAPQSPKAPTAE 263
DB 138 GREHEAPLRGATPDPTDRLSPRPAPQPRRRH-----GRWRPSAS 178
QY 264 ETAAAAAQAQAAAAAAKEREKREORRKORDEGLHLTLTLQCAEVNADNLDAHQTL 323
DB 179 STSDSGSSSSSSSSSSSSSDE---DEDDDG-----NDADHARE---216
QY 324 EIAELATPPTGISTQRYAATFAEAMSAFLVSSCLGLYAPLPGPS-----AAARL- 372
DB 217 -----ARAVGRGPSSAAPAPGR-----TPPPGPPPLSEAPKPRAAATP 258
QY 373 -----HGRVAAAFQVFNGISPFVKFSHTANO-----AIOEAFEREERVHI 413
DB 259 AASAGRIERRRARAAGVAGDATG-----RFTAGQPRRVELDADATSGAFYARYRDGY 310
QY 414 IDDDIMOGLOWPGLFHLASRGGPPRVRLTGLGASMEAL-----EATGKRL 460
DB 311 V-----SGEPWPG-----AGPPPGRVLYGLGDSRPGMGAPAEAEARRFEASGAPA 359
QY 461 SDFADTLG-LPEFCAVAEKAQNVDPKLGVTREAAVAVHMLH-----SLYD 507
DB 360 AWAAPELGDAQAQYALITRLTYTPDAEAMG-----WLONPVVPGDVALDQACFR 409
QY 508 VTGSDSNTLMLIORLAPKVVYTWVEODLSHSGSFLA--RFEVAIHYYSA-----553
DB 410 ISGAARNS-----SSFITGSVAVAPVPHLYGMAAGRFGMGLAHAAAVAMSRRYDRAQ 462
QY 554 ---LFDSDIASYGEDSPERHVEQQLSREIRNVL-AVGCPARTGD-----XKFGSWE 603
DB 463 KGFLLTSLRRYA-----PLLAARENAALTGAAGSPGAGADDEGVAANAAPGE 511
QY 604 KLAQSGFAASLAGS-----AAQAS 624
DB 512 RAVPAGYGAGILALGRLSAPAS 536

```

```

RESULT 14
US-08-843-659-2
; Sequence 2, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopaldi, Rosario
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICPI AS
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee

```

```

; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; APPLICATION NUMBER: US/08/843,659
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARSB:519
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1298 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-843-659-2

```

```

Query Match      4.4%; Score 153.5; DB 4; Length 1298;
Best Local Similarity 20.4%; Pred. No. 0.0007;
Matches 140; Conservative 69; Mismatches 209; Indels 267; Gaps 30;

```

```

QY 53 KRKPAS--DMDLPPRRHVTGDLSDVTAANAAGVSGSPSSASQALPALTQHLQLPRA 110
DB 6 KORGSGPTDGPPTPPSPDRD-----ERGALGKMETEGGDDP-----45
QY 111 FOHNAPEVDVHPAPAAHQAQGEATSTTAWVDGIIRDIISSGGAVSITQLIHNV 170
DB 46 --DHDPD-----HPHDDDARRDGRAPAACTD-----AGEDAGDVSPPQLAL---86
QY 171 EIIHPCNPGIASLLELRSLAADPAFLP-----PPQ 204
DB 87 -----LASVVEEAVRTIPPTDPAASPPRTPAFRADDGDEYDDAADAAGRAPAR 137
QY 205 -POOHALLHGAPAAAGLTPPPPLPDKRRHEHPPRCQQQQQEEBPHAPQSPKAPTAE 263
DB 138 GREHEAPLRGATPDPTDRLSPRPAPQPRRRH-----GRWRPSAS 178
QY 264 ETAAAAAQAQAAAAAAKEREKREORRKORDEGLHLTLTLQCAEVNADNLDAHQTL 323
DB 179 STSDSGSSSSSSSSSSSSSDE---DEDDDG-----NDADHARE---216
QY 324 EIAELATPPTGISTQRYAATFAEAMSAFLVSSCLGLYAPLPGPS-----AAARL- 372
DB 217 -----ARAVGRGPSSAAPAPGR-----TPPPGPPPLSEAPKPRAAATP 258
QY 373 -----HGRVAAAFQVFNGISPFVKFSHTANO-----AIOEAFEREERVHI 413
DB 259 AASAGRIERRRARAAGVAGDATG-----RFTAGQPRRVELDADATSGAFYARYRDGY 310
QY 414 IDDDIMOGLOWPGLFHLASRGGPPRVRLTGLGASMEAL-----EATGKRL 460
DB 311 V-----SGEPWPG-----AGPPPGRVLYGLGDSRPGMGAPAEAEARRFEASGAPA 359
QY 461 SDFADTLG-LPEFCAVAEKAQNVDPKLGVTREAAVAVHMLH-----SLYD 507
DB 360 AWAAPELGDAQAQYALITRLTYTPDAEAMG-----WLONPVVPGDVALDQACFR 409
QY 508 VTGSDSNTLMLIORLAPKVVYTWVEODLSHSGSFLA--RFEVAIHYYSA-----553
DB 410 ISGAARNS-----SSFITGSVAVAPVPHLYGMAAGRFGMGLAHAAAVAMSRRYDRAQ 462

```

QY 554 ---LFSDLSAYGSDSPRRHVEQQLSREIRNVL-AYGCPARTGD-----VKFGSWRE 603
 Db 463 KGFLLTSLRAYA-----PLTARENAALTGAGSPGAGADDEGVAAVAAAAAPGE 511
 QY 604 KLAOSGFRASLAGS-----AAAOAS 624
 Db 512 RAVPAGYGAGGIIAALGRLSAPRAS 536

RESULT 15

US-08-205-506A-2
 ; Sequence 2, Application US/08205506A
 ; Patent No. 5545563
 ; GENERAL INFORMATION:
 ; APPLICANT: Darlington, Gretchen J.
 ; APPLICANT: Wilson, Deborah R.
 ; APPLICANT: Wilde, Margaret
 ; TITLE OF INVENTION: THE HUMAN C/EBP GENE AND VECTORS FOR
 ; TITLE OF INVENTION: ITS EXPRESSION
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fulbright & Jaworski L.L.P.
 ; STREET: 801 Pennsylvania Avenue, NW
 ; CITY: Washington, D.C.
 ; STATE:
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/205,506A
 ; FILING DATE: March 4, 1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Davis, Peter J.
 ; REGISTRATION NUMBER: 36,119
 ; REFERENCE/DOCKET NUMBER: 311.027
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 662-0200
 ; TELEFAX: (202) 662-4643
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 360 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; US-08-205-506A-2

Query Match 4.3%; Score 151; DB 1; Length 360;

Best Local Similarity 21.2%; Pred. No. 0.00019;
 Matches 93; Conservative 35; Mismatches 135; Indels 176; Gaps 19;

QY 6 PPPPLPYRCRCPPLH-----PPPPSSP----- 30
 Db 11 PRPPMSHLQ--SPRAHPSSAAGFPRGAGPAPAPAPAPAPPEPLGICHEHETSIDISAYI 68
 QY 31 -----NHFLHYLHQLDHOEAAAAAAYRKR---PASMDLPPRRRHVDTGLSDVTAATAAA 82
 Db 69 DPAAPNDELLADLFQHSRQOEKAKAAVPTGGGGGDFDP-----GAPA 113
 QY 83 GVGSGAPSSASQALPALPTQLHLPAPFQHNAREVDVPAHRAAQAAGGATATA 142
 Db 114 GPGGAVMPGGA-----HGPPPGY-----GCAAA--G 137

QY 143 WVDGIIRDIIGSSGGANVSTOLHNVRBIHPCNPGIASLLELRRLAADPAAPLPP 202
 Db 138 YLDGRLEPLRYEVGAPALRPVIVKQEPRE-----EDEAKQLALAGLFYQOPPPPP 189
 QY 203 PQQOHALLHGAPA--AAP-----AGITLPPPPPLPDKRHHHPPCQ 243
 Db 190 SHPHRP--HPPPAHLAAPHLOFQIAHCQTTMHLQPDHPPTPPVPV----- 235
 QY 244 QOOOEPRHAPQSPKA-----PTAEETAAAAAAQAAAAAARKEKEORRKO 291
 Db 236 -----SPHPAPALGAAGLPGPSALKGLGAHPDLRASGSGAGAKKSVDKNSNEYVR 290
 QY 292 RDEGHLHLITLLLOCAEAVNADNLDHAQ---TLETAELATPFGTGTQVAAVFAEAM 347
 Db 291 RERNNI-----AVRKSQ-DKAKQNVETQOKVLELTSNDRLRKR-----EQL 333
 QY 348 SARIVSSCLGYAPLPQS 366
 Db 334 S-RELDITLGRITROLPESS 351

Search completed: April 22, 2002, 19:33:11
 Job time: 5561 sec


```

Db 309 ANKLLLEISQLSTPYGTSNORVAAYFSEAMSARLLNSCLGIYAALP--SRMMPQTHSLKM 366
OY 377 AAAFQVFNGISPFVKESHFTANQAIQEAEREERVHIIDIDIMOGLOMPGLFHILLASRPG 436
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 VSAFQVFNGISPLVKFSHTANQAIQEAEREKESVHIIDIDIMOGLOMPGLFHILLASRPG 426
OY 437 GPPRVRLTGLGASMEALATGKRLSDPDTGLPEEFCAVAEKAGNDPEKLGVTREAV 496
    ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 427 GPPHVRLTGLTSMELQATGKRLSDPDTGLPEEFCAVAEKAGNDPEKLGVTREAV 486
OY 497 AVHMLHSLYDVTGSDNTMLIQRLAPKVVYVVEQDLSHSGSEFLARFEVAIHYYSALFD 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AVHMLHSLYDVTGSDNTMLIQRLAPKVVYVVEQDLSHSGSEFLARFEVAIHYYSALFD 546
OY 557 SLDSYGEDESPERRHYVEQQLSREIRNVLAAGPARTGVKESGSRREKLAQSGFRASLA 616
    || |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 547 SLGASVGESEERHYVEQQLSKETRNVLAVGSPRSSEVGFESRKEMQCGFKGISLA 606
OY 617 GSAAQAQSLLLGMFPSPDGYTLVEENGALKLGKDKLCLLTASAMP 661
    :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 GNAATQATLLGMFPSPDGYTLVDNGTLKLGKDKLSLTASAMP 651

```

RESULT 2

```

TS1244
SCARCCROW protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C:Accession: TS1244
R:di Lauro, L.; Wyszocka-Diller, J.; Malamy, J.E.; Pysk, L.; Helariutta, Y.; Freshour
submitted to the EMBL Data Library, July 1996
A:Reference number: 225338
A:Accession: TS1244
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1 653 <DIL>
A:Cross-references: EMBL:U62798; PIDN:AAH06318.1
C:Genetics:
A:Gene: SCARCCROW1
A:Introns: 511/2

```

```

Query Match 53.1%, Score 1861; DB 2; Length 653;
Best Local Similarity 57.3%, Pred. No. 6.9e-102;
Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

```

```

OY 4 PPPPRL-----TPYCRRCRPHLPSPSSRNHLHLHQLDQEAAMAVRRKR 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 PPPHSPLRTTSSGSSSSNNRCPP--PPPP-----LVWVRKRL 48
OY 57 ASDM-----DLPPRR--H-VTGDLSDT-----AAAAGVGGGAPS---SASAQ 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49 ASEMSNPDYNNSSRRPRVSHILDSNNYVTPQOPPSLTAAAYVSSQPNPPLVCGFSG 108
OY 97 LPALPTQLHQLPRAFOHNAPEVDVPAHPAANAAGGEATATTAAYVGIIRDIIGSSG 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 LPVFPSS-----DRGRVNMMSVQPM-----ODSSSSASPTVWVDAILRIHSS- 154
OY 157 GAAVSTIQLIHNVREIHPNCGLASLLELRLLSLADPA-----PLP----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 -TSVSIPLQIQNVRIIFPCNPNIGALLEYRKRSLMLDPSSSSDPSQTFEPPLYQISNN 213
OY 201 --PPQPOOHALLHGAPAAAGLILPPPLPDKRIHNEHPPCQQOQOEPRHPAPQSPK 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 PSPPQQOQOHO-----QQOQOQKPPPPPIQOQERE-----NSSTD 248
OY 259 APTAETAAAA--AAQAAAAAAKERKEQRRKORDEEGCHLTLTLQCAEAVNADLND 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 APPQETATVAVQVTAALREKKEIRKOKODEGLHLLTLTQCAEAVADNLEE 308
OY 318 AHQTLLEIAELATPGCTSTORVAAFAEAMASRLVSSGLIYAPLPSPSPAALHNG-RV 376
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 309 ANKLLLEISQLSTPYGTSNORVAAYFSEAMSARLLNSCLGIYAALP--SRMMPQTHSLKM 366

```

```

OY 377 AAAFQVFNGISPFVKESHFTANQAIQEAEREERVHIIDIDIMOGLOMPGLFHILLASRPG 436
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 VSAFQVFNGISPLVKFSHTANQAIQEAEREKESVHIIDIDIMOGLOMPGLFHILLASRPG 426
OY 437 GPPRVRLTGLGASMEALATGKRLSDPDTGLPEEFCAVAEKAGNDPEKLGVTREAV 496
    ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 427 GPPHVRLTGLTSMELQATGKRLSDPDTGLPEEFCAVAEKAGNDPEKLGVTREAV 486
OY 497 AVHMLHSLYDVTGSDNTMLIQRLAPKVVYVVEQDLSHSGSEFLARFEVAIHYYSALFD 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AVHMLHSLYDVTGSDNTMLIQRLAPKVVYVVEQDLSHSGSEFLARFEVAIHYYSALFD 546
OY 557 SLDSYGEDESPERRHYVEQQLSREIRNVLAAGPARTGVKESGSRREKLAQSGFRASLA 616
    || |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 547 SLGASVGESEERHYVEQQLSKETRNVLAVGSPRSSEVGFESRKEMQCGFKGISLA 606
OY 617 GSAAQAQSLLLGMFPSPDGYTLVEENGALKLGKDKLCLLTASAMP 661
    :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 GNAATQATLLGMFPSPDGYTLVDNGTLKLGKDKLSLTASAMP 651

```

RESULT 3

```

G96688
hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96688
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chn, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matil, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G96688
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE005173; NID:q10092507; PIDN:AGI2907.1; GSPDB:GN00141
C:Genetics:
A:Gene: T27F4.10
A:Map position: 1

```

```

Query Match 17.2%, Score 601; DB 2; Length 511;
Best Local Similarity 34.6%, Pred. No. 4.2e-28;
Matches 146; Conservative 70; Mismatches 150; Indels 56; Gaps 10;

```

```

OY 273 QAAAAAAKERKEQRRKOR-----DEGLHLLTLTLQCAEAVNAD 313
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 RAIPGSAYVRDHYVTRRSKRTRIESELSTRSVVLDSEQGVRLVAALLCAEAVOON 166
OY 314 NLDAHQTLLEIAELATPGCTSTORVAAFAEAMASRLVSSGLIYAPLPSPSPAARLH 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 167 NLKLADALVKNVGLLASSQAGAMRKAYTFABELARIT-----RTY 208
OY 374 GRVAAAFQVFNGL-----SPFVKESHFTANQAIQEAEREERVHIIDIDIMOGLOMP 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 PRDVALLSFSDPLQIHFPESCYLEKFAHFTANQAILEFATAEKVHYVIDLGINCLQMP 268
OY 426 GLFHILLASRPGGPPRVRLTGLGASMEALATGKRLSDPDTGLPEEFCAVA--EKAAND 484
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 ALIQAALLRNPGPPRRLTGLIGSLTDIOEVGKKLQQLASTIGVNEFEKSIALLNLSDLK 328
OY 485 PEKLGV--TREAAVAVHMLH--HSLYDVTGSDNTMLIQRLAPKVVYVVEQDLSHSGS-F 540
    ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 329 PEMLDTRPGLESVAVSVFELHRLLAHPSIDKFLSTISIRPDIMTVVEQDANNGITVF 388

```

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 19:34:35 ; Search time 71.83 Seconds

(without alignments)
340.973 Million cell updates/sec

Title: US-09-265-585c-96

Perfect score: 3503

Sequence: 1 MPPPPPPPLTPYCRRCPP.....KDLCLLTASAMRPVPCR 668

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	6.2	2715	1 TRX2_HUMAN	Q9UM66 homo sapien
2	201	5.7	1048	1 SRA4_RAT	063627 rattus norv
3	199.5	5.7	290	1 TRX2_MOUSE	008550 mus musculu
4	198	5.7	1157	1 SRA4_HUMAN	095104 homo sapien
5	193	5.5	428	1 FXB2_MOUSE	064733 mus musculu
6	187	5.3	1183	1 DRPL_RAT	P54258 rattus norv
7	187	5.3	3421	1 TEGU_HAYEB	P28955 equine herp
8	183.5	5.2	707	1 SEPO_HUMAN	P23246 homo sapien
9	179.5	5.1	1185	1 DRPL_HUMAN	P54259 homo sapien
10	178	5.1	721	1 YK82_MCTU	P10690 mycobacteri
11	171	4.9	620	1 EXTN_TOBAC	P13983 nicotiana t
12	169.5	4.8	1733	1 VNDA_PRYKA	P33485 pseudohomy
13	168.5	4.8	633	1 LA17_YEAST	012446 saccharomy
14	167.5	4.8	1248	1 DIAL_HUMAN	060610 homo sapien
15	167.5	4.8	1790	1 SEPA_HUMAN	P78621 emeritella
16	167.5	4.8	2142	1 BAT2_HUMAN	P48634 homo sapien
17	165.5	4.7	1083	1 T2D3_HUMAN	P00268 homo sapien
18	164.5	4.7	306	1 EXTN_DAVCA	P06599 daucus caro
19	164.5	4.7	901	1 Y298_MOUSE	062523 mus musculu
20	164.5	4.7	564	1 ZYX_MOUSE	015016 homo sapien
21	163.5	4.7	572	1 ZYX_HUMAN	015942 homo sapien
22	162.5	4.6	497	1 WAS2_HUMAN	Q9Y665 homo sapien
23	162	4.6	426	1 EXLP_TOBAC	003211 nicotiana t
24	162	4.6	2205	1 POLN_RUBVT	P13889 rubella vir
25	161	4.6	502	1 WAS3_HUMAN	Q9UY66 homo sapien
26	161	4.6	982	1 CBLE_HUMAN	Q13191 homo sapien
27	161	4.6	1132	1 BAT3_HUMAN	P46379 homo sapien
28	160.5	4.6	339	1 CSP_PLABE	P06915 plasmodium
29	160.5	4.6	1603	1 PSC_DROME	P35820 drosophila
30	160	4.6	347	1 CSP_PLABA	P23093 plasmodium
31	160	4.6	1794	1 YAVI_SCHPO	Q10172 schizosacch
32	159.5	4.6	505	1 WAS1_HUMAN	Q00401 homo sapien
33	158.5	4.5	559	1 WAS1_HUMAN	Q92558 homo sapien

ALIGNMENTS

RESULT 1

TRX2_HUMAN	STANDARD;	PRT; 2715 AA.	
ID	TRX2_HUMAN		
AC	Q9UM66; Q9UM66; Q9Y669; Q9Y668; Q15022;		
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DE	TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN).		
GN	TRX2 OR HRX2 OR MLL2 OR KIA0340.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (LONG ISOFORM).		
RA	Angred P.O., Valvatne H., Jeannoulin F., Adamson A.,		
RA	van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,		
RA	Lamerdin J., Chambon P., Losson R., Stewart A., Asland R.,		
RT	"Mammalian trithorax- and ASH1-like proteins: putative chromatin		
RT	regulators which contain PHD fingers and SET domains."		
RL	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A. (LONG ISOFORM).		
RA	Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhart-Schultz K.,		
RA	Garcia E., Kyle A., Ramirez M., Stillwagen S., Ganes J., Dangnan L.,		
RA	Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,		
RA	Carrano A.V.;		
RT	"Sequence analysis of a 1 Mb region in human 19q13.1";		
RT	Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).		
RC	TISSUE=Brain;		
RA	MEDLINE-97349984; PubMed-9205841;		
RA	Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,		
RA	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. VII.		
RT	The complete sequences of 100 new cDNA clones from brain which can		
RT	code for large proteins in vitro."		
RL	DNA Res. 4:141-150(1997).		
RN	[4]		
RP	SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).		
RC	TISSUE=Testis; and Leukocyte;		
RA	MEDLINE-20105772; PubMed-10637508;		
RA	Hudnune D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,		
RA	Wiedemann L.M., Aparicio S., Caldas C.;		
RT	"MLL2, the second human homolog of the Drosophila trithorax gene, maps		
RT	to 19q13.1 and is amplified in solid tumor cell lines."		
RL	Oncogene 18:7975-7984(1999).		
RN	[5]		
RP	PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).		
RC	TISSUE=Placenta; and Bone marrow;		
RA	MEDLINE-99339983; PubMed-10409430;		
RA	Fitzgerald K.T., Diaz M.O.;		
RT	"MLL2: A new mammalian member of the trx/MLL family of genes."		
RT	Genomics 59:187-192(1999).		
CC	-!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.		

CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
CC	ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
CC	SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL
CC	BLOOD LYMPHOCYTES, AND PLACENTA.
CC	-1- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.
CC	-1- SIMILARITY: BELONGS TO THE PITHORAX FAMILY OF TRANSCRIPTION FACTORS.
CC	-1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC	-1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL, AJ007041; CAB5385.1; -
DR	EMBL, AD000671; -, NOT_ANNOTATED.CDS.
DR	EMBL, AB002302; BAA20763.2; -
DR	EMBL, AF186605; AAD56420.1; -
DR	EMBL, AF104938; AAD17932.1; -
DR	EMBL, AF105279; AAD26113.1; -
DR	EMBL, AF105280; AAD26112.1; -
DR	InterPro: IPR000637; AT-hook.
DR	InterPro: IPR003889; FYRICH_C.
DR	InterPro: IPR003888; FYRICH_N.
DR	InterPro: IPR001965; PHD.
DR	InterPro: IPR002965; P-rich_extensions.
DR	InterPro: IPR001214; SET.
DR	InterPro: IPR002857; ZnF-CXXC.
DR	InterPro: IPR001841; ZnF-ring.
DR	Pfam: PF00628; PHD; 3.
DR	Pfam: PF00856; SET; 1.
DR	Pfam: PF02008; zf-CXXC; 1.
DR	SMART: SM00384; AT_hook; 1.
DR	SMART: SM00542; FYRC; 1.
DR	SMART: SM00541; FYRN; 1.
DR	SMART: SM00249; PHD; 4.
DR	SMART: SM00508; PostSET; 1.
DR	SMART: SM00184; RING; 1.
DR	SMART: SM00317; SET; 1.
DR	PROSITE: PS50280; SET; 1.
KW	DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
KW	Transcription Regulation; Alternative splicing.
FT	DNA_BIND 37 44 A.T HOOK (BY SIMILARITY).
FT	DNA_BIND 110 117 A.T HOOK (BY SIMILARITY).
FT	DNA_BIND 357 365 A.T HOOK (BY SIMILARITY).
FT	ZN_FING 1203 1252 PHD-TYPE 1.
FT	ZN_FING 1253 1303 PHD-TYPE 2.
FT	ZN_FING 1337 1396 PHD-TYPE 3.
FT	DOMAIN 1449 1471 BROMODOMAIN (DIVERGENT).
FT	DOMAIN 2586 2715 SET.
FT	DOMAIN 26 37 POLY-GLY.
FT	DOMAIN 248 255 POLY-PRO.
FT	DOMAIN 362 398 ASP/GLU-RICH (ACIDIC).
FT	DOMAIN 402 771 PRO-RICH.
FT	DOMAIN 808 812 POLY-GLN.
FT	DOMAIN 1863 1870 POLY-PRO.
FT	DOMAIN 2251 2259 POLY-PRO.
FT	VARSPLIC 532 582 VSARSRYIKTPRRMEDPPKPKREVSVYLRLPITTSPP
FT	VSQEPAPVPS -> PLSOSLLPMTLIOISLSIGOWAATTS
FT	ACIDSPMSPLLRLRCPLTGQL (IN ISOFORM
FT	TRUNCATED).
FT	MISING (IN ISOFORM TRUNCATED).
FT	K -> E (IN REF. 5).
FT	S -> Y (IN REF. 5).
FT	CONFLICT 941 941

FT	CONFLICT	1317	1317	E -> O (IN REF. 5).
FT	CONFLICT	1362	1362	H -> Y (IN REF. 5).
FT	CONFLICT	1438	1438	D -> N (IN REF. 5).
FT	CONFLICT	2622	2622	D -> H (IN REF. 5).
SO	SEQUENCE	2715	AA: 293511 MW: C0615B981BBEB7BF	CRC64:

Query Match	6.2%	Score 217:	DB 1:	Length 2715:
Best Local Similarity	23.5%	Pred. No. 0.00032:		
Matches 129:	Conservative 35:	Mismatches 187:	Indels 198:	Gaps 27:

QY	1	MPRP-----PPRLPYRCRCRPHLPRLPPSPSNHFLHYLHLDHOEAAAAA	MYR 53
DB	401	LPRLPPAPSPRLPPRLPPSTSPRLPPCLPPPP-----	434
QY	54	KRPASDML-----PPRRHYVTDLSQVTAANAAGVGGSGAPSSASQAQALPOLTOLHOLPRA	110
DB	435	--VPSRPPLPSPRPAPPAOEQSESPPVVYATCSKKRCP-----PLTQSQRERPA	485
QY	111	FQHNAPEDVANHAPAAHNAQAAGEATASTAWMDGILIRDIIGSSGAASITOLINHV	170
DB	486	--RAGRETSF--PIPIPSTATGCPEDSP-----VAKSTFELKNIR	525
QY	171	ELIHPCNGLASLEFLRLSLADPRLP-----PPQOQHALLHAP	215
DB	526	QFIMVVSARSSRYIKTFRRMDEDP--PKPKVEVSPVLRPPITTSPPVQE---PAP	579
QY	216	AAAPGLTLP--PRLPDKR-----HEPPPCOOQOEHPHAPQSPKA	259
DB	580	VPSRPRAATPSTPRLPEKRRSLIRETFMTSLTRLPP--PRAPPRPAPSPRA	636
QY	260	P-TAEETAAAAAQAQAAAA-----AKERKEQRKORDEEGHLTL	303
DB	637	PATSSRRRLRLRAPQFTPSEAHKLYESVLTLPRLGAEAPDEPPEPADDSP-----	688
QY	304	LQCAE---AVNADLDAHOTLLEIAELAT-----PQGT-----STORVAAYFA	344
DB	689	---AEPERANGRN-----HLSLRFALVYTTPYKAEVSHGAPALNSGPTQAOLOPL	741
QY	345	EAMSARLVSSCLGIYAP--LPPGSP-----AAARLHGVAARFOVNGISPPYKESHFT	396
DB	742	QALQTLQPLPALPPPOLOPPSPQOMPLEKARIAG-----GSLPL-----	785
QY	397	ANQAQOE---AFERERHYIIDLDIMQGLQWPGFLHILASRPGCPRVRLTGLGASMEAL	453
DB	786	--SGVEEMFSLKRAKQLEKIDQOQOK-----VAASPLSP-----GGQMEV	829
QY	454	EATGRKLS	462
DB	830	AGAVKQISD	838

RESULT 2	SRA4_RAT	STANDARD:	PRT: 1048 AA.
AC	O63627:		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	CTD-BINDING SR-LIKE (PROTEIN R44 (FRAGMENT)).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=Hippocampus;		
RX	MEDLINE=96293459; PubMed=8692929;		
RA	Yuryev A., Patturajan M., Litington Y., Joshi R.V., Gentile C.,		
RA	Gehara M., Corden J.L.;		
RT	"The C-terminal domain of the largest subunit of RNA polymerase II		
RT	interacts with a novel set of serine/arginine-rich proteins.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).		

```
CC -1- FUNCTION: MAY ACT TO PHYSICALLY AND FUNCTIONALLY LINK
CC TRANSCRIPTION AND PRE-MRNA PROCESSING (BY SIMILARITY).
CC -1- SUBUNIT: INTERACT WITH THE REPETITIVE C-TERMINAL DOMAIN (CTD) OF
CC RNA POLYMERASE II (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U49058; AAC52660.1; -.
CC InterPro: IPR000504; RRM.
CC Pfam: PF00076; rrm: 1.
CC SMART: SM00360; RRM: 1.
CC PROSITE: PS0102; RRM_1.
CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
CC RNA-binding.
CC NON_TER 1 1
CC DOMAIN 3 6 POLY-PRO.
CC DOMAIN 155 158 POLY-PRO.
CC DOMAIN 422 496 RNA-BINDING (RRM).
CC DOMAIN 630 638 POLY-PRO.
CC DOMAIN 655 661 POLY-PRO.
CC DOMAIN 834 857 POLY-GLN.
CC SEQUENCE 1048 AA; 113701 MW; FSCFBFD309D683E CRC64;

Query Match 5.7%; Score 201; DB 1; Length 1048;
Best Local Similarity 23.0%; Pred. No. 0.00081;
Matches 80; Conservative 31; Mismatches 99; Indels 138; Gaps 16;

QY 1 MPPPPPPPLTPYC-----RCPPPHLP-----PPPPSPNHFLLHYLQLDHQEAAAAMV 52
DB 629 VPPPPPPPP--PFWGPGFNPMLPPGFLPPGPPPTP----- 664
QY 53 KRRASDMDLPPR-----RHYTGDLSDVTAAGVSGAPSSASQDLPALPTQLH 105
DB 665 ---EVS---IPPHHTPSTISPNLYSGARGNAESADSAKMGSGAP-----PAAPTSL- 710
QY 106 QLPRAFGHNAEDVPAHPARAHAQAQGEATASTAMVDGIIRDIIGSSGGAAVSTQL 165
DB 711 PTPPVYTG-----FVSLDGTQGVAPGP-----VIGLQAPSTGLLG-- 745
QY 166 IHNHREIHPNGLASLELRSLAADAPALP-----PPQQOHALLHGAPA 216
DB 746 -----RPGLLPDLR-----PCGMPPHLLRFPMPRPMPPIHMHKGRPP 784
QY 217 AAPAGLTLPP-----LPD-----KRRHEHPPCCQQQ 246
DB 785 PGPGFAPRRHGKGRPPRGPRVRRGGMGLGPGRGSGEDRGRQQQPPQQQQQQQ 844
QY 247 QDEHPAPQSPKAPTAETATAAAAAAQAQAAKKEKEDRRKQORDE 294
DB 845 QQQQQQQQQQQQQPPQSQSQTQQQPPAPSQQPPAQQQPPQDFRNDNRQQ 892

RESULT 3
TRX2_MOUSE STANDARD; PRT; 290 AA.
ID TRX2_MOUSE
AC O08550;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRITHORAX HOMOLOG 2 (MW DOMAIN BINDING PROTEIN 7) (FRAGMENT).
GN TRX2 OR WBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
```

```
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97315177; PubMed=9171351;
RA Bedford M.T., Chan D.C., Leder P.;
RT "FBR W domains and the Abi SH3 domain bind to a specific class of
RT proline-rich ligands."
RL EMBO J. 16:2376-2383(1997).
CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U92455; AAC53192.1; -.
CC MGD: MG1:109565; Wbp7.
CC InterPro: IPR000637; AT_hook.
CC Pfam: PF02178; AT_hook: 1.
CC SMART: SM00384; AT_hook: 1.
CC DNA-binding; Nuclear protein; Transcription regulation.
CC NON_TER 1 1
CC DNA_BIND 1 7 A.T HOOK (BY SIMILARITY).
CC DOMAIN 4 35 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 45 77 POLY-PRO.
CC DOMAIN 253 263 POLY-PRO.
CC NON_TER 280 290
CC SEQUENCE 290 AA; 31202 MW; C49B5565FPC8D2DF CRC64;

Query Match 5.7%; Score 199.5; DB 1; Length 290;
Best Local Similarity 25.4%; Pred. No. 0.00023;
Matches 75; Conservative 22; Mismatches 91; Indels 107; Gaps 12;

QY 2 PPPPPPPPLTPYCRRCPPPHLP-----PPPPSPNHFLLHYLQLDHQEAAAAMVKKRPSDMD 61
DB 46 PVPSPPPPL-----PPSTSPPPASP----- 67
QY 62 LPPRRHNVIGDLSDVTAAGVSGAPSSASQDLPALPTQLHLP-----AFQHHH 115
DB 68 LPP-----VSPPLSLPPYPA--PEKQESSPLVPATGCRKGR 106
QY 116 PEYDVPAHPARAHAQAQGEATASTAMVDGIIRDIIGSSGGAAVSTQLIHNHREIHP 175
DB 107 PPL-TPSQARERAAASGEGTISPSPNPTTSGSLEDSPVIVKSTFLKNIQFIMP 165
QY 176 CNEGLASLELRSLAADAPALP-----PPQQOHALLHGAPAAAPAG 221
DB 166 VVASRSSRVYIKTPRRRMEDEP--PKPKVEASIVRPVATSPAPDE-----FVVPSSPR 219
QY 222 LTLPP--PPPLDPKRR-----HEHPPCCQQQQQDEHPAPQSPKAP 260
DB 220 VPPPSTPVLPDKRSILREPTFRWTSILRELPP-----PPAPPPAPSPAP 269

RESULT 4
SRA4_HUMAN STANDARD; PRT; 1157 AA.
ID SRA4_HUMAN
AC Q95104; Q9UPM1; Q9UPB8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CTD-BINDING SR-LIKE PROTEIN R44 (FRAGMENT).
GN KIAA1172.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
```

[1]
 RP SEQUENCE FROM N.A.
 RA Tassone F., Villard L., Gardiner K.;
 RT "Sequence, genomic organization and map localization of the human SR
 RL protein gene RA4.";
 RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shitani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Mioshima S., Shimizu N., Nordsiek G., Horstischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Borzym K., Gardiner K., Rieseemann L., Dagand E.,
 RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehnach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RN Nature 405:311-319(2000).
 [3]
 RP SEQUENCE OF 195-866 FROM N.A.
 RA TISSUE-Testis;
 RC Ottenwelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,
 RA Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE OF 207-1157 FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=20039618; PubMed=10574461;
 RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
 RA Ohara O.;
 RT "Characterization of cDNA clones selected by the Genemark analysis
 RL from size-fractionated cDNA libraries from human brain.";
 RN DNA Res. 6:329-336(1999).
 CC -1- FUNCTION: MAY ACT TO PHYSICALLY AND FUNCTIONALLY LINK
 CC TRANSCRIPTION AND PRE-MRNA PROCESSING (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACT WITH THE REPEATING C-TERMINAL DOMAIN (CTD) OF
 CC RNA POLYMERASE II (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PROBABLY PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF023142; AAD09327.1; -
 DR EMBL: AF001711; -; NOT_ANNOTATED_CDS.
 DR EMBL: AL117417; CAB55911.1; -
 DR EMBL: AB032998; BAA86486.1; -
 DR InterPro: IPR002965; P_r10ch_extensn.
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm.1.
 DR PRINTS: PRO1217; PRICHEXTENSN.
 DR SMART: SM00360; RRM.1.
 DR PROSITE: PS50102; RRM.1.
 DR PROSITE: PS00030; RRM_RNP.1; FALSE_NEG.
 KW RNA-binding; Alternative splicing.
 KW NON_TER 1
 FT DOMAIN 1
 FT DOMAIN 5 8
 FT DOMAIN 165 168 POLY-ALA.
 FT DOMAIN 314 321 POLY-PRO.
 FT DOMAIN 323 326 POLY-ALA.
 FT DOMAIN 518 592 POLY-PRO.
 FT DOMAIN 726 733 RNA-BINDING (RRM).
 RN DOMAIN

FT DOMAIN 750 756 POLY-PRO.
 FT DOMAIN 967 976 POLY-GLN.
 FT VARSPIC 776 798
 FT CONFLICT 270 274
 FT CONFLICT 293 293
 FT CONFLICT 296 296
 FT CONFLICT 309 309
 FT CONFLICT 312 312
 FT CONFLICT 562 562
 FT CONFLICT 570 570
 FT CONFLICT 581 581
 FT CONFLICT 608 610
 FT CONFLICT 614 616
 FT CONFLICT 622 622
 FT CONFLICT 626 626
 FT CONFLICT 818 836
 FT CONFLICT 837 866
 FT CONFLICT 854 854
 FT CONFLICT 1149 1149
 FT CONFLICT 1157 AA: 126846 MW: 176A778996BFB01B CRC64;
 SQ
 Query Match 5.7%; Score 198; DB 1; Length 1157;
 Best Local Similarity 22.6%; Pred. No. 0.0013;
 Matches 86; Conservative 24; Mismatches 98; Indels 172; Gaps 17;
 QY 1 MPP-----PPPPPLPYCR-----RCPPPLP-----PPSSPNFLHYLDHDA 46
 DB 716 IPPGFGGVPPPPPPFLRPGFNPMHLRPGFLRPPPPITP----- 759
 QY 47 AAAAMVKKRPASMDLPP-----RHHVGDLS-----DYTAAGAAGV 84
 DB 760 -----FVS-----IPPTPTISIPNSTIAGINEDTKDISIGNPIPTVVGARGNAE 807
 QY 85 GGSAGPSSASQALPALTOLHOLPFAFONHAEVDVRAHAPARAQAOGGEATASTAVY 144
 DB 808 SSGSVAMYGSAVPAAPATNL-PPPTVQ-----FVS----- 837
 QY 145 DGLIRIIGSSGAIVSTIQLHNVREIINPCNGLASILERLNSLAADRA--PL--- 199
 DB 838 -----LLTGQGA-----PGVILQLAPSTGLGARPGLIPLQRP 872
 QY 200 -----PP-----PPQQHALLHGAPAPAGULTRP----- 227
 DB 873 PCMPRHLDRLMPRRPMHMRGPRGSGFAMRRHGMKGRPPHGFVRPGMP 932
 QY 228 -----PLPDKRRHHRPPCQOQOQDEPHAPQSPKAPAEETAAAAAQAQAAAAA 279
 DB 933 GLGPGPGGGRGDRGRRQRPQRPQOQ-----PQRAQRPQOQOQOQRPSPSQRPPTVQ 988
 QY 280 AKERKEQRKKO---RDEE 295
 DB 989 OPQGFRRNDNRQOFNSGRDQ 1008
 RESULT 5
 EXB2_MOUSE
 ID EXB2_MOUSE STANDARD; PRT; 428 AA.
 AC 064733;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 35, Last annotation update)
 DE FORKHEAD BOX PROTEIN B2 (TRANSCRIPTION FACTOR FKX-4).
 GN FOXB2 OR FKX4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=97014266; PubMed=8661101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
  in the central nervous system.";
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
  Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
  regulated.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92591; CAA6335.1; -
DR EMBL; X71942; CAA50744.1; -
DR TRANSFAC; T02442; -
DR MGD; MG11347468; Foxd2.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head.1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH.1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
DR KEGG; K04400; Fork_head; Transcription regulation.
DR DNAStrand; Nucleic protein; Transcription regulation.
FT DOMAIN 12 103 FORK-HEAD.
FT DOMAIN 139 153 POLY-HIS.
FT DOMAIN 156 162 POLY-HIS.
FT DOMAIN 163 172 POLY-PRO.
FT DOMAIN 217 231 POLY-ALA.
FT DOMAIN 249 258 POLY-ALA.
FT DOMAIN 321 330 POLY-ALA.
FT DOMAIN 396 399 POLY-ALA.
SQ SEQUENCE 428 AA; 45170 MW; DB8A8EF01E94AB10 CRC64;

Query Match 5.5%; Score 193; DB 1; Length 428;
Best Local Similarity 26.8%; Pred. No. 0.0008;
Matches 86; Conservative 30; Mismatches 93; Indels 112; Gaps 17;

```

```

OY 256 SPKAPTAETAAAAAQA 276
  ||| | ||| | : : :
DB 400 SEPTABLEPTAAGRADSKGS 420

RESULT 6
DRPL_RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLUTSIAN ATROPHY PROTEIN).
GN DRPLA.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., La S.-H., Schilling G.,
  Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
  homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
  dentatorubral-pallidolutsian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
  MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
  (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31777; AAA80337.1; -
DR EMBL; X89453; CAA61623.1; -
DR InterPro; IPR002951; Atrophin.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-SER.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADCF9B1F CRC64;

Query Match 5.3%; Score 187; DB 1; Length 1183;
Best Local Similarity 19.6%; Pred. No. 0.0052;

```

Matches 118; Conservative 30; Mismatches 154; Indels 300; Gaps 24;

```

Oy 2 PPPPPP-----PLTPYCRRC---PPPHLP-----25
    |||||
Db 165 PPSPPPPDSIPRQPSGSEPRHPSVPTGYIHAMEPPISRLPQGGAPRPHPOLYPSGA 224
Oy 26 -----PPSSNHFLLHYLQHDQEAAMAMVKKRPASDMDLPPRRHVTGDSLDTA 78
    |||||
Db 225 GGGVLSPGMPG-----KGGAAASSVCPRSSGKQHP-----TTP 260
Oy 79 AAAAGVGGSGAPSSASADLPALPTQLHQLPP--AFQHHAREVDV-----121
    |||||
Db 261 IPISSSGASGAPRKPRTPVGAGNLPSPAPPTGPHVTPLPPRPLRLNNASASBP 320
Oy 122 -----AHPAPAHQAQAG-----EAT 137
    |||||
Db 321 MCAOPITPGHLPSPHAMGCGMGLPPRPCKPTLAPSRPLRPASSABGPPMRYPYSSCS 380
Oy 138 ASTTAMVDGIIRDIISSGGAASVITQLIHNVREIHPNCNGLASLELRSLADPA 197
    |||||
Db 381 SSSVA-----ASSSSAATS-----QYRASQTLPSY-----PH 408
Oy 198 PLPPRPQ-----POQHALLHGAAPAAAGLLPRLPP-----LPDKRH--236
    |||||
Db 409 SFPPTSMVSNNQPKYTPQSLPSQAVWSQGP-----PPRPYGRLLPNNNTHPG 459
Oy 237 -----EHP-PC-----QOQOQEEHPAPQ-----SPKAPTAETAAAAAQA 274
    |||||
Db 460 PEPPTGGSTAHRRPARHNNHQOQOQRPQRPQOHHNGNSGRRPCA-----507
Oy 275 AAAAAAKRKEQRKKQKDEBGLHLTLTLLOCAEVNANDIDAHQTLLEIATLPFGT 334
    |||||
Db 508 -----YRPHLESSNHNH-----PYN-524
Oy 335 STORVAAYFAEAMARSCLGLYAPLPGPSAAARLHGVAAPQVENGISPFVFSH 394
    |||||
Db 525 -----MSPSLGSLRPYPPPAHLPPSHGQVYSTQAGPMG--PVSSSS 565
Oy 395 FTANQAIQEAFAFEREVRHIIIDIMQGLQWGLFPHILASRGGPPRYRLTGLGASMEALE 454
    |||||
Db 566 NSSGSSSQAAVSCS---HPSSSGPQGASYP-----PFPVPPITTSATLSTVI 611
Oy 455 AT 456
    ||
Db 612 AT 613

```

RESULT 7
TEGU_HSVB STANDARD: PRT: 3421 AA.

AC P28935;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE LARGE TEGUMENT PROTEIN.
GN 24.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae;
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RA "The DNA sequence of equine herpesvirus-1.";
RC Virology 189:304-316(1992).
RL -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HSV-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86664; AAB02459.1; -
DR PIR: G36797; WZBBE6.
DR HSSP: P04002; 1WFA.
SQ SEQUENCE 3421 AA; 367078 MW; 5075EFE4739BB7AC CRC64;

Query Match 5.3%; Score 187; DB 1; Length 3421;
Best Local Similarity 26.7%; Pred. No. 0.017;
Matches 82; Conservative 32; Mismatches 119; Indels 74; Gaps 13;

```

Oy 1 MPPPPPPITPYCRRCPPHLP PPPSSPNHFLHYLQHDQEAAMAMVKKRPASDM 60
    |||||
Db 2570 LPAPAPLPQSTSKAASGATQSDGKTLTLDVPTQSKDKVP-----LPQSTSKAASG-----2607
Oy 61 DLPPRRHVTGDSLDTAAAGVGGSGAPSSASADLPALPTQLHQLPPAFQHHAREVDV 120
    |||||
Db 2608 --PPPTLPPAPPLPQSTSKAASGATQSDGKTLTLDVPTQSKDKVP-----VPPTDK 2659
    |||||
Oy 121 PAHPAPAHQAQAGGATATTAWDGIIRDIISS-----SGGAASVITQLIHNVREI 174
    |||||
Db 2660 PSTTTPAALKQS--DASKPPTAIIQH--QKIGTPTPKDSQDKPTDNASAPVGSVTP 2715
Oy 175 PCNPGLASLELRSLADPAFLPRLPP-----OQOHAL--LHGA-----PAAAPAG 221
    |||||
Db 2716 DGTFG-----AKRPPKDAVPDDTKQPVKRSLSQVCGRRYIRPSIGPK 2761
Oy 222 LTLPPPLPDKRHRHPPPP-----CQOQOQEEHPAPQSKAPTAETAA-AAAAAQA 277
    |||||
Db 2762 FTGPPGTYTP---VHGLPPSDSNVTQSTKEPRKPAVETPAAPASAAAAPAAAPAKSAAA 2818
Oy 278 AAKERK 284
    ||
Db 2819 PAAAPAK 2825

```

RESULT 8
SFPO_HUMAN STANDARD: PRT: 707 AA.

AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPLICING FACTOR, PROLINE AND GLUTAMINE-RICH (POLYRIMIDINE TRACT-
DE BINDING PROTEIN-ASSOCIATED SPLICING FACTOR) (PTB-ASSOCIATED SPLICING
DE FACTOR) (PSF) (DNA-BINDING P52/P100 COMPLEX, 100 KDA SUBUNIT).
GN SFPO OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RA MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RA "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RT Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RA MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RA "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1b5 monoclonal antibody.";
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.


```

RT "Novel triplet repeat containing genes in human brain: cloning,
RL expression, and length polymorphisms."
RL Genomics 16:572-579(1993).
CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC Ovary, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
CC LIVER, THYMUS AND LEUKOCYTES.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC PALLIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUTS/BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: D31840; BAA06626.1; -.
DR EMBL: U23851; AAB50276.1; -.
DR EMBL: L10377; -: NOT_ANNOTATED_CDS.
DR HSSP: P00651; ILRA.
DR MIM: 125370; -.
DR InterPro: IPR002951; Atrophin.
DR PRINTS: PR01222; ATROPHIN.
KW Triplet repeat expansion; Polymorphism.
FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).
FT 302 305 POLY-PRO.
FT 376 382 POLY-SER.
FT 386 397 POLY-SER.
FT 442 447 POLY-PRO.
FT 479 483 POLY-HIS.
FT 484 497 POLY-GLN.
FT 504 507 POLY-PRO.
FT 564 574 POLY-SER.
FT 704 707 POLY-PRO.
FT 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT 94 94 MISSING (IN REF. 2).
FT 333 333 Y -> H (IN REF. 2).
FT 339 339 M -> I (IN REF. 2).
FT 541 541 P -> T (IN REF. 3).
FT 1028 1028 G -> A (IN REF. 2).
SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

Query Match 5.1%; Score 179.5; DB 1; Length 1185;
Best Local Similarity 21.5%; Pred. No. 0.013;
Matches 166; Conservative 69; Mismatches 259; Indels 279; Gaps 37;

QY 2 PPPPP-----PRLPYCRRCPP-----H 21
DB 442 PPPPPRYLLANSNANRPPRPSTGAOSTAHNPVSTNNHHNQQOQQOQQOQNNHNG 501
QY 22 LPPPPSSPNHFL-----LHYLDHLDHQAAMAAWVRKRPASDMLPRPRRHVT----- 70
DB 502 SGPPPGAFRRPLRGGGSHNHN--PYANSPFGLSLRPRREGRAHLPRPHSOVSISQAGPN 559
QY 71 -----GDLSDYTAAGG-----VGSAGD-----SSAQAQPLPFTQLHQL 107
DB 560 GPPVSSSSSSSSSTSGSYPSCSHPSPOGPGAGAPYPPRPVTVTTSATLSTVATAYASS 619
QY 108 PPAFHNAPEVDVPAHRAAANAQAAGBATASTTAWDGIIRDIIGSSGAAVSIQOLI 167

```

```

DB 620 PAKYKASP-----PCPPYGRAPSGAKKATTPPY-----KQSPSPFTGPP 666
QY 168 NWEIILPCNPGLASLELRSLSLAADPAPLPPPPQOQHLLHGAPAAAPAGL-TLPP 226
DB 667 GYKGTSPGPGTFK-----PGSPYCGPLP-----AGBSGLSLPP 705
QY 227 PRLPDKRRHNPPOCOQOQOQEEP--HPAQSPKAPT----- 261
DB 706 PPAAP-----ASGPPLSATQIKQEPAEYETPSYPVPAASPPPKVVDPSHASQARF 761
QY 262 -----AETAAAAAQAQAAAAAKERKEQORRORDEE 295
DB 762 NKHLDFNSCARSDLYFVPLEGSKLAKKRADLVKVRREARQAREREKEREREKE 821
QY 296 -----GHLTLTLQCAEA-VNADNLDDA-HQTLLELAELATPF--GTSQRYAAYFA 344
DB 822 RERERELEERSVKYLAQEGRAPECPSLGPVPHR-----PPREGSAVAIVPLIG 872
QY 345 -EAMSAFLVSSCIGLYA--PLPPGSPAARLH-----GRVAAAFQVFGNISPFVKFSH 394
DB 873 PDRPALRTISE--YARPHVMSPGN--RNHPFVPLGAVDPGLGYN--VPALYSSD 922
QY 395 FTANQAIQFAFERER-----VHIIDLMQGLQMPGLEHT-----LASRPGGPRPV 441
DB 923 PAARERERARERDLDRLLKPGFEVYKPSLEPLHGVPGGLPPFRHGGLAQGPSP-- 979
QY 442 RLGLGASMEALEATGKRLLSDFADTLGLPFECAVAEKAGN-----DPEKLGVTARRA 495
DB 980 -----GLHFPFPHPSLG-PLEKERLALAGPLRDMKSAEFLAERQHA 1023
QY 496 VAVHMLHSLYDVTGSDSNTLMLQRLARVVTWVEQDLSHSGSL-ARFVAIHYSA- 553
DB 1024 ERVAGL-----GNDP-----LARTQLMNTVPHHGHSHISHLHLHQDPAHAASAS 1070
QY 554 ---LFPSTL-DASVGEDSP-----EKHHVVEQDLSREIRNYVA 586
DB 1071 VHPDLPLASGSHLRIPRPACTLPNPLLNPHLENEVLRHQLFAAPYRDLP 1123

RESULT 10
YK82_MCTTU STANDARD: PRT: 721 AA.
AC Q10690;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 73.6 KDA PROTEIN RV2082.
GN RV2082 OR MCTY49.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=96342230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jajels K., Krogan A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: SOME, TO M. TUBERCULOSIS RV0029.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: 273966; CAA98194.1; -
 KW Tuberculosis; RV2082; -
 DR Hypothetical protein; Complete proteome.
 FT DOMAIN 295 307 ALA-RICH.
 SQ SEQUENCE 721 AA; 73564 MW; 28CAC21029EDDC57 CRC64;

Query Match 5.1%; Score 178; DB 1; Length 721;
 Best Local Similarity 24.7%; Pred. No. 0.0091;
 Matches 96; Conservative 20; Mismatches 125; Indels 148; Gaps 17;

QY 3 PPPPPPLTPYCRRCPPPHLPP-----PPSSPNHFLHYHQDHEAAAAAVRRK 55
 DB 250 PPTAPPTSPGTAHPMPGPPPGTVSPPLPSAP-----AVGCGPSV 293
 QY 56 PASDMDLPPRRHYTGDSLDTAAAGVG-GSGAPSSASAC--LPALPTQLHQLPPAF 111
 DB 294 PAAGM--PAAAAATAPLSQSLGQFTTGMTGTAPAAAGALSGALHAATEPLPP- 350
 QY 112 QHAPPEVDVAHPAPAAHAQAGEATATTAWVDGITRDIIGSSGAAVSTOLLHNRE 171
 DB 351 --APPTPTPTVTPV-----ATATVA-----GI----- 372
 QY 172 ITHPCNGLSLLELRSLAADPAFLPPPOQHA-----LHGAPAAPAGLTLP 225
 DB 373 ---PHIDSA-----TPSPAPIAPPTDMSAMPTIAPMANGPAPAPAPAA 419
 QY 226 PPPPLP---DKRRHEPPCQOQOEHP-----APQSPKA----- 259
 DB 420 PAGLPAYGDLR-----PVTPPATPTPTGTPTSGAAVTSPSPAAGSLMSPVVKST 474
 QY 260 -----PRAETAAAAAQAQAAAAAKEKEQRKQ-----RDEGHLILT 301
 DB 475 APATTAQPSNPPTPLASATAATGTGAAGDTSRAAEQRLRLIDTVAROEGLSW- 532
 QY 302 LLQCAEAVNANDNDHQTLLLEIAELAT 330
 DB 533 -----AAGLRNGQTTTLVTLDAS 551

RESULT 11
 EXTN_TOBAC
 ID EXTN_TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HGGPMT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 RT glycoprotein gene in lateral root initiation.";
 RU genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN

CC GLYCOSTYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X13885; CAA32090.1; -
 DR PIR: S06733; S06733.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN ? 620
 FT REPEAT 70 73 EXTENSIN.
 FT REPEAT 148 151 H-A-P-P.
 FT REPEAT 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 242
 FT REPEAT 236 242
 FT REPEAT 205 242
 FT DOMAIN 205 620
 FT DOMAIN 499 600 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 4.9%; Score 171; DB 1; Length 620;
 Best Local Similarity 26.3%; Pred. No. 0.018;
 Matches 72; Conservative 10; Mismatches 92; Indels 100; Gaps 12;

QY 2 PPPP---PPPLTPYCRRCPPPHLPPPPSSPNHFLHYHQDHEAAAAAVRRKRA 58
 DB 348 PPPPPYSPPPP--PSYSPPTPTLPPPPSSP-----PPPSFSPPTTY 389
 QY 59 DMDLPPRRHYTGDSLDTAAAGVGSGAPSSAS-----AQLPALPTQLHQLPPAFQ 112
 DB 390 EOSPPPPPAY-----SPPLAPPTYSPPPTTSPPTTYAQPPPLPPTYSPPPAY 441
 QY 113 HNAPEVDVA---HPAPAAHAQAGEATATTAWVDGITRDIIGSSGAAVSTOLLHN 168
 DB 442 PPPPPTYSPPPTYSPPPAYAQ----- 465
 QY 169 VREIITHPCNPGIASLLELRSLAADPAFLPPPOQOHALHAPAAAPAGLTLP 228
 DB 466 -----PPPPPTYS-----PPPAYSPPPSP--IYSPPOVOPLPTFSPP 506
 QY 229 PLDKRHEHPPCQOQOEHPA--POSPKAPT 261
 DB 507 P---RRIHLPPP--HKQRPPTPTTYGDPSPPT 535

RESULT 12
 VNDA_PVKA
 ID VNDA_PVKA STANDARD; PRT; 1733 AA.
 AC P33485;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE PROBABLE NUCLEAR ANTIGEN.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=33703;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021039; PubMed=2171211;
 RA Vieck C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
 RT oriented open reading frame: characterization of their promoter and
 RT enhancer regions.";
 RU Virology 179:365-377(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M34651; AAA47471.1; -
 DR PIR; B45344; B45344.
 FT DOMAIN 112 117 POLY-THR.
 FT DOMAIN 179 1733 GLY-RICH.
 FT DOMAIN 192 196 POLY-SER.
 FT DOMAIN 271 298 POLY-PRO.
 FT DOMAIN 304 308 POLY-ARG.
 FT DOMAIN 883 889 POLY-GLY.
 FT DOMAIN 1398 1405 POLY-GLY.
 SQ SEQUENCE 1733 AA; 172166 MW; 0C8CDBBE475B5E2 CRC64;

Query Match 4.8%; Score 169.5; DB 1; Length 1733;
 Best Local Similarity 22.0%; Pred. No. 0.07;
 Matches 109; Conservative 27; Mismatches 153; Indels 207; Gaps 20;

OY 2 PPPPPPLP-----TP-YCRGCPPHLP-----PPSSPNHFLHYLHQLDHOE 45
 DB 70 PPHPPRPPODHHPTPARDHDPDHDLPRTTRDHOHRRPPTTTTINKDPHPODPL 129
 OY 46 AAAAMVKKRP---ASDDLPPRRHVTGDISDTAAAAAGVSGGAPSSAS----- 94
 DB 130 LPTKLTQEDPHLPRTDPPSAKTHNHOD-----PPGGCPSTSHHHNHODPP 178
 OY 95 ----AQLPALP---TQLHQLPRAFOHNAPEVDVPAHPA---AHAAQAGGATASTTA 142
 DB 179 GGGPPSPRRPSTSSSHGPPSTRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPP 237
 OY 143 WWDGIIRLIIGSSGAASITQLIHNVREIT-HCENPGLASLLERLSLLAA-----D 195
 DB 238 -----FHSSEKLFSPHGEGEGE-----DRGTAGGEGGDRD 270
 OY 196 PAPPPLPPOOHALLHGAAPAAAGLTLPPRPLPDKRHHNPPPCOOQOEHPHPAPQ 255
 DB 271 PPPPPPPRRPP-----PLPPPPPP-----PP-----QPPRAGG 301
 OY 256 SP-----KAPTAETAAAAAQAQAAAAAAKERKEORRORDEE- 295
 DB 302 SARRRRRGGCPRRGRRGGRRRAEGTEAAADAEEEDGDDEDEDRAEGEGRED 361
 OY 296 -----GILHLLTLLLCQA 309
 DB 362 GGEGRGAGGAGESESESGRAEGAPSAEQOVAGVGLLVVRDGLHLDGPERAANGPA 421
 OY 310 VNADNLDDAHOTLLEIATLPTGSTQRYAAYFAEAMASARLVSSCLGLVAPLPDPSPA 369
 DB 422 VAAADADDLH-----RVPVLAGAGRGAGRGPRGVLHG--APGGAD 461
 OY 370 ARL-----HGRVAA 378
 DB 462 AGLEGKVPHEGRGA 477

RESULT 13
 LA17_YEAST STANDARD; PRT; 633 AA.
 AC Q12446;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROLINE-RICH PROTEIN LASI7.
 GN LASI7 OR YOR181W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Toth-E.A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO S.POME WSP1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D78487; BAA11386.1; -
 DR EMBL; 275089; CAA9390.1; -
 DR SGD; S0005707; LAS17.
 DR InterPro; IPR000697; RanBPL_WASP.
 DR InterPro; IPR001960; WH1.
 DR InterPro; IPR003124; WH2.
 DR Pfam; PF00568; WH1; 1.
 DR Pfam; PF02205; WH2; 1.
 DR SMART; SM00461; WH1; 1.
 DR SMART; SM00246; WH2; 1.
 FT DOMAIN 185 190 POLY-PRO.
 FT DOMAIN 323 329 POLY-PRO.
 FT DOMAIN 342 348 POLY-PRO.
 FT DOMAIN 352 358 POLY-PRO.
 FT DOMAIN 385 391 POLY-PRO.
 FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 448835563AA2645 CRC64;

Query Match 4.8%; Score 168.5; DB 1; Length 633;
 Best Local Similarity 24.5%; Pred. No. 0.026;
 Matches 91; Conservative 36; Mismatches 143; Indels 101; Gaps 18;

OY 3 PPPPPPLTPYCR-----RCPPHLPPPPPSPNHFLHYLHQLDHOEAAAAAMVKKRPAS 58
 DB 323 PVPPPPPMRTTTEGSGVRLP--APPPP-----RRGAP 354
 OY 59 DMDLPPRRHVTGDLSDVTAAGAAAGVSGGAPSSASAOULPALPTOLHQLPRAFOHNAPEV 118
 DB 355 P---PPPHRVVTSNTLN-----SAG-GNSLLPQATGRRGPAPPP-----PPRASRPTPV 400
 OY 119 DVAPHAPAAHA-QAGGATAST-----TAMWDGIIRLIIGSSGAASVITQLIHNV 169
 DB 401 TMOQNPOQIINNSNRPPGYOTNSMSSPPPPVTTFTTLTPONTAAGORAVLPQ---NT 457
 OY 170 RELIHPNGLASLELRSLIADRAPLPPRPOQHALLHGAAPAAAGLTLPPPP 229
 DB 458 QA-----PSQATNVV-----APPPPASLGSGQITQSAVSA-----PIPT 494
 OY 230 LPDKRHHNPP---CQOOQOEHPHPAPQSPKAPTAETAAAAAQAQAAAAAAKERKEE 286
 DB 495 LPSTTSAPPPPAFLVTLQPOSGAGAPPPPPOMPATSTSGGSPAETTGADAG---RDA 550
 OY 287 QRKKQKDEGSHLLTLLLOCAEAVNANDNDHQLTLEI-AELAP-----FGSTQRYA 340
 DB 551 LASTAGAGGIGAL-----RKVDKSQLDKPSVLQEARGSASPPAAAGNGTGGGP 603
 OY 341 AYFAEAMARL 351
 DB 604 ASLADALAAAL 614

```

RESULT 14
DIAPH_HUMAN STANDARD: PRT; 1248 AA.
AC 060610; 09UC76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 1 (DIAPHANOUS-RELATED FORMIN 1) (DRFL).
GN DIAPH1 OR DIAPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98028756; PubMed=9360932;
RA Lynch E.D., Lee M.K., Morrow J.E., Welch P.L., Leon P.E., King M.-C.;
RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
RL homolog of the Drosophila gene diaphanous."
RL Science 278:1315-1318(1997).
RN [2]
RP SEQUENCE OF 218-817 FROM N.A.
RX TISSUE-Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Wakamabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Niinomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 727-765 AND 1121-1145.
RX TISSUE=platelet;
RA Reinhard M., Gleich K., Abel K., Haffner C., Jarichau T., Hoppe V.,
RA Jockusch B.M., Walter U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
RL ligand for profilins."
RL EMBO J. 14:1583-1589(1995).
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DER PROTEIN COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
CC ACTIN POLYMERIZATION IN HAIR CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.
CC -1- DOMAIN: DRES ARE REGULATED BY INTRAMOLECULAR GBD-DAD INTERACTION
CC RHO-GTP ACTIVATES THE DRES BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT
CC NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRE AUTOREGULATORY DOMAIN (DAD).
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
CC TO INCLUDE INTRONIC SEQUENCE.
CC -1- DATABASE: NAME=Hereditary hearing loss homepage;
CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF051782; AAC05373.1; -.
DR EMBL; AK023345; BAB14533.1; ALT_SEQ.
DR MIM; 602121; -.
DR MIM; 124900; -.
DR InterPro: IPR003104; FH2.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
KW Coiled coil; Repeat; Deafness.
FT DOMAIN 63 260 GBD.
FT DOMAIN 157 457 FH3.
FT DOMAIN 460 563 COILED COIL (POTENTIAL).
FT DOMAIN 588 743 FH1 (PRO-RICH).
FT DOMAIN 748 1190 FH2.
FT DOMAIN 1015 1172 COILED COIL (POTENTIAL).
FT DOMAIN 1173 1187 DAD.
FT DOMAIN 1189 1192 ARG/LYS-RICH (BASIC).
FT DOMAIN 804 804 T -> TSRA (IN REF. 2).
FT CONFLICT 1132 1133 RK -> AE (IN REF. 3).
SQ SEQUENCE 1248 AA; 138978 MW; ED1F514/CF9A886 CRC64;

Query Match 4.8%; Score 167.5; DB 1; Length 1248;
Best local Similarity 19.6%; Pred. No. 0.062;
Matches 114; Conservative 53; Mismatches 171; Indels 243; Gaps 25;

QY 2 PPPPPPPPLPYCRKPPPLP-----PPSSPNHLLHYHQLDQEAAMAAVRRKR 55
DB 600 PPPPPPP-----PPPLPGTAISPPPLS----- 625
QY 56 PASDMLPPRRHVGTGDSVYATAAAGVSGSPASASAPL--ALPTQLHQLPFAQ 112
DB 626 --GDATIPPP-----PPLP-----GVGIPSS--LPGTAIP-----PPPL 660
QY 113 HNAEVDVPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPA 172
DB 661 PGSAIPPPPPPLP-----GSAG----- 678
QY 173 IHCNGLASLELRSLAADPAPLPPPOOHALLGHPAHPAHPAHPAHPAHPAHP 232
DB 679 IPPPPPLPG-----EAGMPPPP-----LPGGPPIPPPPPPG 713
QY 233 KRHEHPPCQQQQQEEP--HPAPSPKAP-----TAETAA 267
DB 714 GRIPIPPPPGKMPPPPPPGFVPAAPVLPGLITRKKLYKPEVQLRRPWKSLVAEDLSQ 773
QY 268 -----AAAAAQAAMAAAKERREQRKQRDEGLHLTL 302
DB 774 DCFWTKKEDRFENNELFAKLITLTFSAQTKRKDQGGKKSVQKKVKE----- 824
QY 303 LQCAAAVAADNL-----DDAHTLEIELELPTFSTQRYAAATFAEAMARL 351
DB 825 -LKVLDSTKAQNLSTFLGFRMPYQEIKNVLEVNE-AVLTESMQLNLIKQNPEDQK 882
QY 352 VSSCGLVAPLPGPS-----AAARLHGRVAAP-----QVFNGISPFYKFSHTAN 398
DB 883 LSLAKDEYDLDASDEFGVGMQTVPRRLRALILFKLQFSQVFN-----IKPIVSVT 937
QY 399 QAIQAEFEEREVHHIIDIDIMOGLOWPGLFHTLASRPGPPRVRLTGLAGSMEALATGK 458
DB 938 AACCEELRKSEFSNLEITLL-----VGYNMAGSRNAGAFNLSFL-CKLRDTKSTQ 991
QY 459 RL-----SDPADTLGPFEECAVAEKAAGVDEKL 488
DB 992 KMTLLHFLAELCENDYPDYLPKPDLAHV-EKASRYSAENL 1031

RESULT 15
SEPA_EMENTI

```

```

ID  SEPA.EMENT STANDARD: PRT: 1790 AA.
AC  p78621: 000760:
DT  20-AUG-2001 (Rel. 40, Created)
DT  20-AUG-2001 (Rel. 40, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  CYTOKINESIS PROTEIN SEPA (FHL/2 PROTEIN) (FORCED EXPRESSION
DE  INHIBITION OF GROWTH A).
GN  SEPA OR FIGA.
OS  Emericella nidulans (Aspergillus nidulans).
OC  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC  Eurotiaceae; Trichocomaceae; Emericella.
OX  NCBI_Taxid=5072;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=97361829; PubMed=9218790;
RA  Harris S.D., Hamer L., Sharpless K.E., Hamer J.E.;
RT  "The Aspergillus nidulans sepa gene encodes an FHL/2 protein involved
RT  in cytokinesis and the maintenance of cellular polarity."
RL  EMBO J. 16:3474-3483(1997).
RN  [2]
RP  REVISTIONS TO 143-153; 207: 1071-1109 AND 1644.
RA  Hamer L., Harris S.D., Sharpless K.E., Hamer J.E.;
RL  Submitted (MAR-2001) to the EMBL/Genbank/DBD databases.
RN  [3]
RP  SEQUENCE OF 559-1790 FROM N.A., AND FUNCTION.
RC  STRAIN=FGSC26;
RX  MEDLINE=95229045; PubMed=7713416;
RA  Marhoul J.F., Adams T.H.;
RT  "Identification of developmental regulatory genes in Aspergillus
RT  nidulans by overexpression."
RL  Genetics 139:537-547(1995).
CC  -1- FUNCTION: INVOLVED IN CYTOKINESIS. OVEREXPRESSION RESULTS IN
CC  GROWTH INHIBITION.
CC  -1- DOMAIN: DRES ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC  RHO-GTP ACTIVATES THE DRES BY DISRUPTING THE GBD-DAD INTERACTION
CC  (BY SIMILARITY).
CC  -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC  -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FHL) DOMAIN.
CC  -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC  -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC  -1- SIMILARITY: CONTAINS 1 DRE AUTOREGULATORY DOMAIN (DAD).
CC  -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. BN11
CC  SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U83658; AAB63335.3; -
DR  EMBL: L36341; AAA33306.1; ALF_SRO.
DR  InterPro: IPR001104; FH2.
DR  Pfam: PF00769; ERM; 1.
DR  SMART: SM00498; FH2; 1.
KW  Cell division; Coiled coil.
FT  DOMAIN 258 486 GBD.
FT  DOMAIN 376 718 FH3.
FT  DOMAIN 724 811 COILED COIL (POTENTIAL).
FT  DOMAIN 955 1136 FHL (PRO-RICH).
FT  DOMAIN 1141 1658 FH2.
FT  DOMAIN 1435 1566 COILED COIL (POTENTIAL).
FT  DOMAIN 1586 1600 DAD.
FT  DOMAIN 1608 1611 ARG/LYS-RICH (BASIC).
FT  CONFLICT 1071 1109
FT  CONFLICT 1476 1476
FT  CONFLICT 1504 1504
FT  CONFLICT 1790 AA: 197355 MW; 192136DE2E2F2A75B CRC64;
SQ  SEQUENCE

```

```

Query Match 4.8%; Score 167.5; DB 1; Length 1790;
Best Local Similarity 18.7%; Pred. No. 0.093;
Matches 123; Conservative 44; Mismatches 192; Indels 299; Gaps 25;

QY 2 PPPPPPLTPYCRRCRCPPHLPPPPSSPNHFLYHLQHLQHOEAAAAMVRRKRPASDMD 61
DB 1017 PPPPPPPRAPHQLSQAAPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1057
QY 62 LPPRRHVTGDLSDVTAANAAGVGGSGAPSSASAOPLPQLQHLPPAFQHHAPVDVPE 121
DB 1058 PPPP-----PGGIGGP-----PP-----P 1071
QY 122 AHPAPAAHQAQGEATASTTAVWDGIINDIIGSSGAAVSITQLIHNVRELIHPCNPGLA 181
DB 1072 PPPPP-----GGFGG----- 1082
QY 182 SLELRSLRLAADPAPLPPPPPOQHALLHGAAPABAGLTLPPPPPLPKRRHHP 241
DB 1083 -----PPPPP-----PPGCGFGPPPP-----PP 1103
QY 242 CQQQQQEEPH-----PAQSPKAPTAETMAAANAQAANAQAANAQAANAQAANAQA 296
DB 1104 -----PGGAFGVPPPPPPPGTVIGGWRANYLASQAPSHALPVMSSIRPKK----- 1150
QY 297 LHLTLTLQCAVANADNLDADHQTLEIAELATPEGTSTORVAAVFAEAMSAARLVSSCL 356
DB 1151 -----LKALHWDKVDTPQVTVW-----ATHGTTPEKEKYYELAKRGVLDVE 1194
QY 357 GYAPLPPGSPAARLHGVAAPQ-----VFNGIS-----PVKSHFTANAOIDEA 404
DB 1195 RLF-----MAKETRIFGGVAAKORKDKKOIISNDLKNFOIALSKFSQEPAEVVR-- 1246
QY 405 FERERVHIIIDIMQGLQWPLFHLIASRPGPPRVRLTIGASMEALVATGKRSLDFA 464
DB 1247 -----RIHCDAILDNM-----VMEFLQD----- 1268
QY 465 DTGLPFEFCAAVAKAGNV-----DPEKLGVTREBAVAHWLHLSLYDVTGSDNST 515
DB 1269 -----EMCTVPENVSKLMAPIYSKDWGCPDAANTEREDQPSELTREDOYLYLTA 1321
QY 516 LMLQLRLAKVYTVWQDLSHSGSFLARVEALIHYSALFQSLDASGEDSPERRHVQ 575
DB 1322 YKAAARRALALTRSEPPDEH--ISAKLREVVAVRSESILRDSVS----- 1362
QY 576 LLSREIRNVLAVGPARTDGVKFSWREKLAQSGFRAASLAGSAAQAASLLGMFSPD 633
DB 1363 -LMNVIGLILDIG-----NFMNDANKQAQ-GFKLSLA-----RLGMYKDD 1401

```

Search completed: April 22, 2002, 19:44:18
Job time: 583 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 19:33:15 ; Search time 122.38 Seconds
(without alignments)
798.414 Million cell updates/sec

Title: US-09-265-585C-96

Perfect score: 3503

Sequence: 1 MPPPPPPPLTPYCRRCPP.....KDLCLLASAMRPVPPCR 668

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3503	100.0	668	10 Q9FUZ7	Q9FUZ7 zea mays (m
2	1865	53.2	653	10 Q9M84	Q9M84 arabidopsis
3	1861	53.1	653	10 Q96304	Q96304 arabidopsis
4	1747	49.9	819	10 Q9AVK4	Q9AVK4 pisum sativ
5	1005.5	28.7	405	10 Q9FH21	Q9FH21 arabidopsis
6	670	19.1	623	10 Q9ST59	Q9ST59 triticum ae
7	666	19.0	630	10 Q9ST48	Q9ST48 zea mays (m
8	635	18.1	625	10 Q9MB96	Q9MB96 oryza sativ
9	616	17.6	547	10 Q9SRP9	Q9SRP9 arabidopsis
10	602	17.2	493	10 Q9AS97	Q9AS97 oryza sativ
11	601	17.2	511	10 Q9C8Y3	Q9C8Y3 arabidopsis
12	594.5	17.2	662	10 Q65367	Q65367 arabidopsis
13	594.5	17.0	587	10 Q23642	Q23642 arabidopsis
14	593.5	16.9	587	10 Q23725	Q23725 arabidopsis
15	590.5	16.9	587	10 Q9SLH3	Q9SLH3 arabidopsis
16	590.5	16.9	532	10 Q23724	Q23724 arabidopsis
17	589.5	16.8	533	10 Q9LQ78	Q9LQ78 arabidopsis
18	589.5	16.8	532	10 Q23643	Q23643 arabidopsis
19	574	16.4	617	10 Q9LW09	Q9LW09 oryza sativ

20	572	16.3	523	10 Q9LFE3	Q9LFE3 arabidopsis
21	570.5	16.3	526	10 Q9LPT0	Q9LPT0 arabidopsis
22	562	16.0	490	10 Q9LDL7	Q9LDL7 arabidopsis
23	561	16.0	658	10 Q9CAN3	Q9CAN3 arabidopsis
24	557	15.9	584	10 Q9FL03	Q9FL03 arabidopsis
25	548	15.6	413	10 Q9S7H5	Q9S7H5 arabidopsis
26	544.5	15.5	593	10 Q9SD03	Q9SD03 arabidopsis
27	538.5	15.4	447	10 Q9ARF6	Q9ARF6 capsella ru
28	536.5	15.3	482	10 Q9LPR8	Q9LPR8 arabidopsis
29	525	15.0	428	10 Q9ZSP2	Q9ZSP2 lycopersico
30	513.5	14.7	445	10 Q9ZWC5	Q9ZWC5 arabidopsis
31	499.5	14.3	306	10 Q9XE53	Q9XE53 arabidopsis
32	498	14.2	352	10 Q9XE51	Q9XE51 arabidopsis
33	470.5	13.4	542	10 Q9SCR0	Q9SCR0 arabidopsis
34	466	13.3	610	10 Q9LRT5	Q9LRT5 arabidopsis
35	458	13.1	1502	10 Q9LNX6	Q9LNX6 arabidopsis
36	457	13.0	483	10 Q9SUF5	Q9SUF5 arabidopsis
37	442	12.6	718	10 Q80933	Q80933 arabidopsis
38	439.5	12.5	583	10 Q9SNB8	Q9SNB8 arabidopsis
39	436	12.4	101	10 Q9AVK3	Q9AVK3 pisum sativ
40	423.5	12.1	375	10 Q23566	Q23566 arabidopsis
41	417.5	11.9	1336	10 Q81074	Q81074 arabidopsis
42	404.5	11.5	287	10 Q9MOM5	Q9MOM5 arabidopsis
43	399.5	11.4	284	10 Q9XE57	Q9XE57 arabidopsis
44	398	11.4	808	10 Q9XE58	Q9XE58 arabidopsis
45	373.5	10.7	582	10 Q9FWN6	Q9FWN6 oryza sativ

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	668 AA.
Q9FUZ7	Q9FUZ7			
AC	Q9FUZ7			
DT	01-MAR-2001 (TREMblrel. 16, Created)			
DT	01-MAR-2001 (TREMblrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMblrel. 16, Last annotation update)			
DE	SCARECROW.			
GN	SCR.			
OS	zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;			
OC	Panicoidae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20407145; PubMed=10948251;			
RA	Lim J., Heierliutta Y., Specht C.D., Jung J., Sims L., Bruce W.B.,			
RA	Diehn S., Benfey P.N.;			
RT	"Molecular analysis of the SCARECROW gene in maize reveals a common			
RT	basis for radial patterning in diverse meristems."			
RL	Plant Cell 12:1307-1318(2000).			
DR	EMBL: AF263457; AAC13663.1; -			
SQ	SEQUENCE 668 AA: 71162 MW: 92874310462B2621 CRC64:			

Query Match	100.0%;	Score 3503;	DB 10;	Length 668;
Best Local Similarity	100.0%;	Pred. NO. 2.8e-222;		
Matches 668;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MPPPPPPPLTPYCRRCPPPHLP	PPPPSSPNHFLHYLHQLDHOEAAAAMVKKRPASDM	60
Db	1	MPPPPPPPLTPYCRRCPPPHLP	PPPPSSPNHFLHYLHQLDHOEAAAAMVKKRPASDM	60
QY	61	DLPPRRHVTGDLSDVTA	AAAGVGSGAPSSASQALPALTQULQLPAPFOHNAREVDY	120
Db	61	DLPPRRHVTGDLSDVTA	AAAGVGSGAPSSASQALPALTQULQLPAPFOHNAREVDY	120
QY	121	PAHPADAAHQAGEATASTTAAV	DDGIIRDITGSSGGAASVTITQLHNARETIHPCNGL	180
Db	121	PAHPADAAHQAGEATASTTAAV	DDGIIRDITGSSGGAASVTITQLHNARETIHPCNGL	180

```

OY 181 ASLELELRSLAADPAAPLPPPOQOHALHGAFAAPAGLTLPPPLPPDKRHHPP 240
DB 181 ASLELELRSLAADPAAPLPPPOQOHALHGAFAAPAGLTLPPPLPPDKRHHPP 240
OY 241 PQQOQOQEEPPHAPSPAPTAETAAAAAQAQAAAAAKEERKORRORDEBGLHL 300
DB 241 PQQOQOQEEPPHAPSPAPTAETAAAAAQAQAAAAAKEERKORRORDEBGLHL 300
OY 301 TLLLOCAEVNANDLDDHOTLLETAELATPGSTORVAAVFAFAMGARLVSSGLIXA 360
DB 301 TLLLOCAEVNANDLDDHOTLLETAELATPGSTORVAAVFAFAMGARLVSSGLIXA 360
OY 361 PLPPGSPAAARLHGRVAAAFQVFNCSIPPVKFSHETANQAIQEAFFEREERVIIDLDIMQ 420
DB 361 PLPPGSPAAARLHGRVAAAFQVFNCSIPPVKFSHETANQAIQEAFFEREERVIIDLDIMQ 420
OY 421 GLQWGLFHIIASRRGGPRRRLTGLGSMALERTGRLSDPADTGLPPEFCVAAEKA 480
DB 421 GLQWGLFHIIASRRGGPRRRLTGLGSMALERTGRLSDPADTGLPPEFCVAAEKA 480
OY 481 GNVDEKIGVTRREAVAAHMLHSLDYVTGSDSNTMLIORLAPKVYTMVEODLSHSGSF 540
DB 481 GNVDEKIGVTRREAVAAHMLHSLDYVTGSDSNTMLIORLAPKVYTMVEODLSHSGSF 540
OY 541 LARFEALHYYSALEFDSLDAASYGEDSPERHVEOQLSREIRNVLAAGPARTGVKFS 600
DB 541 LARFEALHYYSALEFDSLDAASYGEDSPERHVEOQLSREIRNVLAAGPARTGVKFS 600
OY 601 WREKLAOSGFRAASLAGSAAQAASLLGMPDSGYTYLVEENGALKLGKDLCLLTASAMR 660
DB 601 WREKLAOSGFRAASLAGSAAQAASLLGMPDSGYTYLVEENGALKLGKDLCLLTASAMR 660
OY 661 PLOVPPCR 668
DB 661 PLOVPPCR 668

```

```

RESULT 2
OY 09M384 PRELIMINARY: PRT: 653 AA.
AC 09M384:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SCARECROWI.
GN F24B22.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloeker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.,
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132957; CAB70996.1;
SQ SEQUENCE 653 AA; 71505 MW; 009A4C48C6DA0616 CRC64;

```

```

Query Match 53.2%; Score 1865; DB 10; Length 653;
Best Local Similarity 57.4%; Pred. No. 1.3e-114;
Matches 405; Conservative 66; Mismatches 122; Indels 112; Gaps 18;
OY 4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHYLQDLHQEAAAAAAVYKRP 56
DB 12 PPHSPPLRTSSSSSSNNNGPPP--PPPP-----LVMYKRL 48
OY 57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAAGVGSGAPS--SASAO 96

```

```

DB 49 ASESSSNDYNNSSRRPRRSHLLDSNNTVTQOPPSLTAATVSSQNPPLSVCGFSG 108
OY 97 LPALPTQHLQLPFAFOHHAEDVDVPAHPAPAAHAQAAGENTASTAWVDGIITDIISSSG 156
DB 109 LPVPPS-----DRGGRNVMSVQPMQ---QSSSSASPTWVAIDIRDLIHS- 154
OY 157 GAAVSTOLHNVREIHPNPGSLASLELRSLAAPRA-----PLP----- 200
DB 155 -TSVSIPLQLQNVNDIIFPCNPMLGALLELRSLMLPLDSSSDSPSPQTFEPYLOISNN 213
OY 201 -PPPOQHALLHGAFAAPAGLTLPPPLPPDKRHHPPPCQOQOQEEPPHAPQSPK 258
DB 214 PSPPOQOQOQO-----QOQOQHPPPPPIQOQERE-----NSSTD 248
OY 259 APTAETAAAAA-AAQAAAAAAAKEERKORRORDEBGLHLITLLLOCAEVNANDLDD 317
DB 249 APPQETVATVAVQNTNAEALREKREIKRQKQDEBGLHLITLLLOCAEVNANDLEE 308
OY 318 AHOTLLEIAELATPGSTORVAAVFAFAMGARLVSSGLIXAPLPPGSPAAARLHG--RV 376
DB 309 ANKLELEISOLSTPYGTSQORVAATPSEAMASARLNSCLGITYALP--SRMPQTHSLKM 366
OY 377 AAQOVFNCSIPPVKFSHETANQAIQEAFFEREERVIIDLDIMQGLQWGLFHIIASRRG 436
DB 367 VSAFQVFNCSIPPVKFSHETANQAIQEAFFEREERVIIDLDIMQGLQWGLFHIIASRRG 426
OY 437 GPRVRLTGLGSMALERTGRLSDPADTGLPPEFCVAAEKAQNVDEKIGVTRREAV 496
DB 427 GPRVRLTGLGSMALERTGRLSDPADTGLPPEFCVAAEKAQNVDEKIGVTRREAV 486
OY 497 AVHMLHSLDYVTGSDSNTMLIORLAPKVYTMVEODLSHSGSFARFEALHYYSALEP 556
DB 487 AVHMLHSLDYVTGSDAHTLMLIORLAPKVYTMVEODLSHSGSFARFEALHYYSALEP 546
OY 557 SLDASYGEDSPERHVEOQLSREIRNVLAAGPARTGVKFSWREKLAOSGFRAASLA 616
DB 547 SLGASGESEERHVEOQLSREIRNVLAAGPARTGVKFSWREKLAOSGFRAASLA 606
OY 617 GSAQAASLLGMPDSGYTYLVEENGALKLGKDLCLLTASAMR 661
DB 607 GNAATQATLLGMPDSGYTYLVDNGTLKLGKDLCLLTASAMR 651

```

```

RESULT 3
OY 096304 PRELIMINARY: PRT: 653 AA.
AC 096304:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE SCARECROWI.
GN SCARECROWI.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Di Laurenzio L., Wysocka-Diller J., Malamy J.E., Pysch L.,
RA Helariutta Y., Freshour G., Hahn M.G., Feldmann K.A., Benfey P.N.;
DR EMBL: U62798; AAB06318.1;
SQ SEQUENCE 653 AA; 71535 MW; 502057F2666B161B CRC64;

```

```

Query Match 53.1%; Score 1861; DB 10; Length 653;
Best Local Similarity 57.3%; Pred. No. 2.3e-114;
Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;
OY 4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHYLQDLHQEAAAAAAVYKRP 56
DB 11 PPHSPPLRTSSSSSSNNNGPPP--PPPP-----LVMYKRL 48
OY 57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAAGVGSGAPS--SASAO 96

```



```
Db 12 PPHSPKLTSSGSSSSNNRGP--PPPP-----LVMVRKL 48
Qy 57 ASDM-----DLPPRR--H-VTGLSDVT-----AAAAGVGSGAPS---SASAQ 96
Db 49 ASEMSSNDYNNSSRPKRYSHLDSNNTVTPOQPSLTAATAVSSQPNPLSVCGSG 108
Qy 97 LPALPTQHLQHPAFQNHAPRVDYPAHPARAHAAGGEMASTTAVNDGIRIITSSG 156
Db 109 LPVPS-----DRGRNVMSVOPMD---ODSSSSASPTVWDALIRDLHSS- 154
Qy 157 GAAVSTOLINHVREIHPNCPGLASLELRSLADPA---PLP----- 200
Db 155 -TSVSTPQLIONVDDIIFPCNPNIGALLEYRLSLMLDSSSDSPQOFPEPLXQISNN 213
Qy 201 --PPPOQOHALHGAPAAAPAGTLPPRPLPDKRRHHPPCQOQOQOEEHPAPQSPK 258
Db 214 PSPPOQOQOHH-----OQOQOHHKPPPPYIQOQERE---NSSTD 248
Qy 259 APTAEPTAAAAA-AAOAAAAAAKEREKRRKORDEEGHLITLLQCAEVNADLDD 317
Db 249 APPOPEVTATVPVAVQNTAEALRERKEEIKRQODEEGHLITLLQCAEASADNLEE 308
Qy 318 AHQTLLEIAELATPEGTSTORVAAVFAEAMASARLVSSCLGLYAPLPGPSAAARLHG-RV 376
Db 309 ANKLLEISQLESTPYGTSAPQVAAVFEAMASARLVNSCLGIYALP--SRMMQOTSLKM 366
Qy 377 AAAQOVENGISPEYKESHFTANQAIQCAFEREERVHIIDIMOGLOMPCGLPHILASRP 436
Db 367 VSAQOVENGISPLVKFSHFANQAIQCAFEREKESVHIIDIMOGLOMPCGLPHILASRP 426
Qy 437 GPRVRLTGLGASMEALATGKRLSDPADTLGLPFEFCVAEKAAGVNDPKLGTREAV 496
Db 427 GPRVRLTGLGTSMEALQATGKRLSDPTDKLGLPFEFCPLAEKAGNDTELRNKEAV 486
Qy 497 AVHMLHSLDYVTGSDSNLMLIQRLAPKYVTVEODLSHSGFLARFEVAIHYSALFD 556
Db 487 AVHMLHSLDYVTGSDSNLMLIQRLAPKYVTVEODLSHSGFLARFEVAIHYSALFD 546
Qy 557 SLDSYGEDSEERHVEQQLSREIRNVAVGAPARTGVKFGSWREKLAQSGFRASLA 616
Db 547 SLGASYESEERHVEQQLSREIRNVAVGAPARTGVKFGSWREKLAQSGFRASLA 606
Qy 617 GSAAGASLLGMFSPDGYTLVEENGALTKGKDCLLTASAMP 661
Db 607 GNAATQATLLIGMFPDGYTLVDNGTLKLGKDCLLTASAMP 651

RESULT 4
Q9AVK4 PRELIMINARY: PRT: 819 AA.
AC Q9AVK4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE SCARECROW GENE REGULATORY-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CV. ALASKA;
RA Sassa N., Matsushita Y., Nakamura T., Nyunoya H.;
RT "Molecular characterization and in situ expression pattern of pea
RT SCARECROW gene."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB048713; BAB39155.1;
SEQUENCE 819 AA; 90372 MW; 41867BD6DC72ADFA CRC64;
```

Query Match 49.9%; Score 1747; DB 10; Length 819;

```
Best Local Similarity 62.9%; Pred. No. 9,5e-107;
Matches 361; Conservative 64; Mismatches 91; Indels 58; Gaps 8;
Qy 135 EATASTTAWVDGIIRIDILSSGGAASITOLINHVREIHPNCPGLASLELRSLAA 194
Db 245 EETSATTNMDGLDKDLHITSN--SVSIPOLINHVREIHPNCPNIALVLEHRLR--LIT 300
Qy 195 DPAPLPPPO---POQHALHGAPAA-----PAGTLPPPPPLD 232
Db 301 EPNTQPEERKRNSTEQGVNNGVNLAAASNVSVKLMNRVDVYPTSLHSDSSTLLN 360
Qy 233 KRHEH-----PPPCQOQOQOEEHPAPQSPK--APTAEPTAA 267
Db 361 QONQNMFPNMGATQINNNNPSVSLVTLPSQPLSTQDQOQOHLQHPEDLAP----AT 415
Qy 268 AAAAAAATAAAKEREKRRKORDEEGHLITLLQCAEVNADNDDAHQTLLEIAE 327
Db 416 TTTTSAELALARKKKEELKEOKKDEBHLITLLQCAEASVNLQAKMLLEISQ 475
Qy 328 LATPEGTSTORVAAVFAEAMASARLVSSCLGLYAPLPGPSAAARLHGRYAAAFQVFN 387
Db 476 LTPPEGTSQORVAAVFEAMASARLVSSCLGIYATLPVSSHTPH--NQKVASAFQVFN 533
Qy 388 PFVKEHFHTANQAIQCAFEREERVHIIDIMOGLOMPCGLPHILASRPGPPVRLTGL 447
Db 534 PFVKEHFHTANQAIQCAFEREERVHIIDIMOGLOMPCGLPHILASRPGPPVRLTGL 593
Qy 448 ASMEALATGKRLSDPADTLGLPFEFCVAEKAAGVNDPKLGTREAVAVMLHSLYD 507
Db 594 TSMETLEATGKRLSDPANKLGLPFEFFPAEKVGNIDVKNLVSSFAVAVMLHSLYD 653
Qy 508 VTGSDSNLMLIQRLAPKYVTVEODLSHSGFLARFEVAIHYSALFDSYGEDSP 567
Db 654 VTGSDSNLMLIQRLAPKYVTVEODLSHSGFLARFEVAIHYSALFDSYGESE 713
Qy 568 ERHVVEQQLSREIRNVAVGAPARTGVKFGSWREKLAQSGFRASLAAGASALL 627
Db 714 ERHVVEQQLSREIRNVAVGAPARTGVKFGSWREKLAQSGFRASLAAGASALL 773
Qy 628 GMFSPDGYTLVEENGALTKGKDCLLTASAMP 661
Db 774 GMFSPDGYTLVDNGTLKLGKDCLLTASAMP 807

RESULT 5
Q9FH21 PRELIMINARY: PRT: 405 AA.
AC Q9FH21;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE SCARECROW GENE REGULATORY-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA MEDLINE-99397451; PubMed-10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL: AB017067; BAB08425.1;
SEQUENCE 405 AA; 44893 MW; 6809D9EEBEBOA8F6 CRC64;
```

Query Match 28.7%; Score 1005.5; DB 10; Length 405;
Best Local Similarity 53.2%; Pred. No. 2.1e-58;

```

Matches 207; Conservative 65; Mismatches 106; Indels 11; Gaps 6;
QY 278 AAABERKE--EQRKKQDEBGLHLLTLLOCAEAVNADNDDAHQTLLLELAELATPGTGS 335
Db 17 SSARRLEEFPEETLENDGAAAIKLLSLLOCAEYVAADHLREASTLLSEISICSPGSS 76
QY 336 TORAAAYFAEMASRLVSSCL-GLYAPLPFGSPAAARLHGVAARLFOVNGISPFFVFSH 394
Db 77 PERVAAYFAQALQTRVISSYISGACSPLE-KPLTVVOSQKIPFALQTYNSVSLIKFSH 135
QY 395 FTANAOIAOEAFEREERHVIDLDIMQGLQWPGFPHILASRPGPPRYRLTGLGASMALE 454
Db 136 FTANOAIFQALDGEDSVHIIIDLAVQGLQWPFALPHILASPRKLRISIRITGFGSSDLA 195
QY 455 ATGRKLSFADTGLPPEFCAVAAKAGN-VDEPKLGTREAVAVHMLHSIVDYQSDS 513
Db 196 STGRRLAFASSLNPPEFPIEGITIGNLIDPSQATRGCAVAVVHMQRKLDVYGNL 255
QY 514 NTLMLIORLAKPVYTWEDLSH--SGSFARFEVAIHYSALPDSLDASYGEDSPERHV 571
Db 256 ETELELRKLKRLNLTIVVQELSTYDDGSGFLGRFVEALHYSALEFDALDGDGESSGERFT 315
QY 572 VEQQLLSREIRNVLAVGPARFTGVKFGSWREKTLAQSGFRAASLAGSAAQAQSLILGMP 631
Db 316 VEOYLGTGEIRNIYAHGSGRR---KRMKKEELSRVGFPRVSLRGNPATQAGILLGMLP 371
QY 632 SDGYTLVEENGALTKGWKDCLLLTASAMR 660
Db 372 WNGYTLVEENGTLRGWKDSLTLTASAMK 400

RESULT 6
Q9ST59 PRELIMINARY: PRT; 623 AA.
AC Q9ST59;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GIBBERELIN RESPONSE MODULATOR.
GN RHT-DIA.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9347734; PubMed=10421366;
RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,
RA Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,
RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
RT "Green revolution" genes encode mutant gibberellin response
RT modulators."
RL Nature 400:256-261(1999).
DR EMBL; AJ242531; CAB5155.1; -.
SQ SEQUENCE 623 AA; 65337 MW; F2BAC34996D0A84F CRC64;

Query Match 19.1%; Score 670; DB 10; Length 623;
Best Local Similarity 30.9%; Pred. No. 4.2e-36;
Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

```

```

QY 228 PRLPDKRRHEHPPPCQOQOQEEBHPAPQSPKAPAEETAAAAAQAQAAAAAKERKEEQ 287
Db 170 DSVYRDKRRKRTGSSSTSSSSSSSSSLGGAKSSVVEAPRVAAAAAATPALRVVV----- 224
QY 288 RKRQDEBGLHLLTLLOCAEAVNADNDDAHQTLLLELAELATPGTSTORVAAVFAAM 347
Db 225 --VDTQEGAIRLVHALLACAEAVQOENLSAALVKQIFLLAASOGAMRKVAAYFGEAL 282
QY 348 SARLVSSCGILAPLPFGS-----PAAARLHGVAARLFOVNGISPFFVFSHTANAOIAOE 403
Db 283 ARRVFR-----FRPODSSLDAAFADLLHAF-----FYESCYLKFAHTAQAQILE 330
QY 404 AFEREERHVIDDIMGQGLQWPGFPHILASRPGPPRYRLTGLA-----SMEALATGKR 459
Db 331 AFAGCNRVHVDGFKQGNQWALLQALALRRGSPFRFLITGVGPPQDEIDALQOYQWK 390
QY 460 LSDPADTGLPPEFCVA-----EKAGNVDEPKLGTREAVAVHMLH--HSL 505
Db 391 LAQFARTIRVDYRGLVATLADLEPFMLQPEGEDPNE---BPEYLAIVSVFEMHRL 446
QY 506 YDVTGSDSNTMLIORLAKPVYTWEDLSH--SGSFARFEVAIHYSALPDSLDASYGE 564
Db 447 LAQPGALEKVLGTVAVRPRIVTVEQENHNSGTFLOFTESLHYSTMPDSLEGSSG 506
QY 565 DSPER-----HVEQQLLSREIRNVLAVGPARFT-GDYKFGSWMREKLAOS 608
Db 507 GGPSEVYSSGAAAPAAAGTDQYMSFVYLGROICNVACGARTERHETLQWRRRLGNA 566
QY 609 GFRAASLAGSAAQAQSLILGMP-SDGYTLVEENGALTKGWKDCLLLTASAMR 660
Db 567 GFETVHLGSNAYKQASTLLALFAGDGYKVEKEGCTLIGWHTRLIATRSAMR 619

RESULT 7
Q9ST48 PRELIMINARY: PRT; 630 AA.
AC Q9ST48;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GIBBERELIN RESPONSE MODULATOR (FRAGMENT).
GN Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9347734; PubMed=10421366;
RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,
RA Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,
RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
RT "Green revolution" genes encode mutant gibberellin response
RT modulators."
RL Nature 400:256-261(1999).
DR EMBL; AJ242530; CAB51557.1; -.
FT NON_TER 630
SQ SEQUENCE 630 AA; 66028 MW; 3D56851726C51042 CRC64;

Query Match 19.0%; Score 666; DB 10; Length 630;
Best Local Similarity 31.9%; Pred. No. 7.7e-36;
Matches 209; Conservative 80; Mismatches 253; Indels 114; Gaps 21;

```



```

Db 179 RLVAALVACAEAIHOENINLADALVKRVGTLGAGAGAMKVATYFAQALARRIYRDYTA 238
OY 355 CGLIYAPLPSPGSPAARLHGVRVAAAFQVFNIGSPVFKSHFTANOAIQEAEREERHII 414
Db 239 EFDVCAAVNPSEFEVLEH-----FYESCPIYKFAHTANOAIIEAATTARRAVVI 289
OY 415 DIDIMOGIOWPGLFPHILASRPGPPRVRLTIG-----ASMALERTGKRLSDPADTLGLP 470
Db 290 DGLNGMQMOPALMOALALRPGPPSRLTIGPQTEINSLSIQGLKLAQFQNMGE 349
OY 471 EEFCAV-AEKANNDPEKLV-TREAVAVHHL--HSLYDYGSDSTLMLIQRLAKV 526
Db 350 EEFKGLAAESLSDLEPEMEFTRPESETLVVNSVFELHRLARSGSEIKLNTVYAIKPSI 409
OY 527 VMVEBODLSHG-SFLARFEVAIHYYSALFSDLSAGDESPERHVEQOLLSREIRVYL 585
Db 410 VAVVEQEANHNIGVFLDRFNEMALHYSSLFSDLEDSTYSLPQDR-VMESEVYLGNQILNV 468
OY 586 AVGPAPRGDVKFGS-WREKLAQSGFRAASLAGSAAQASILLGNFSPS-DGYTLVEENGA 643
Db 469 AAEGSDRVERHETAAQMRIRKMSAGFDPIHLGSSAFKQASMLSLYATGDCGYRVEENDGC 528
OY 644 LKLGKDCILITASAMR 660
Db 529 LMTGMOTRPLTTSAWK 545

```

RESULT 10

```

O9AS97 PRELIMINARY: PRT: 493 AA.
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DE PUTATIVE OSGAL.
GN P0707D10.30.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPOBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RA "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:p0707D10."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AP002910, BAB40172.1; to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE 493 AA; 51864 MW; 5FD8670EB8899492 CRC64;

```

Query Match 17.2%; Score 602; DB 10; Length 493;
 Best local similarity 35.2%; Pred. No. 9, 1e-37;
 Matches 171; Conservative 61; Mismatches 192; Indels 62; Gaps 16;

```

OY 195 DPAPLPPOQOQHALLHGAPAAVAGTLPPRPPLDPKRHHNPPPCQOQOQDEPPHAP 254
Db 2 DTFEPQWMDP-----AASSGLDAGFLPPAAVAPDDGYGYPD-----PAG 43
OY 255 OSPKAPTAETRNA--AAAAAAMAAAKERKEORRQKOREEGJHLTLLOCAEVNA 312
Db 44 ADVDAALPERFAAPPCAPDAAAVALMRREE-----EVAIGIRLVHLLMSCAGIEA 97
OY 313 DN-----LDDAHOITLEIETLPTGSTQORVAAVFAEAMSRALVSSCGLTAPLP 365
Db 98 GDHALASQLADSHALAAVS-AASGIG---RVAVHTTALSRLLP-----PSPVA 144
OY 366 SPAARLHGVRVAAAFQVFNIGSPVFKSHFTANOAIQEAEREERHIIIDIMOGIOWP 425
Db 145 PPTTDAEH--AFLYHNHYEACPYLKFAHFTANOAILLEAFHGCCHVHYIDPSLMOGIOWP 201
OY 426 GLFHILASRPGPPRVRLTIG-----ASMALERTGKRLSDPADTLGLPFEFCVAAEKA- 480

```

```

Db 202 ALIQALALRPGCPPELITIGIPSPPTGREDLADVGLRLADLARSVRFESFGVAANSL 261
OY 481 GNDPPEKLVTKREAVA---VHMLHSLYDVTSSD--SNLMLIQRLAKVYTMVQDLS 535
Db 262 DEVRPMMLOIAPGEVAVNSVQLHRLIGDPADQAPIDAVLDCVAVSRIFIFVIEQED 321
OY 536 HSGS-FLARFEVAIHYYSALFSDLSAGDESPERHVEQOLLSREIRVLAAGPA-RT 593
Db 322 HNTGFLDRTEALFYTSANFSDLSAASGAGNMAE-AYLQREICDVCGEAARE 380
OY 594 GDVKFGSWREKLAQSGFRAASLAGSAAQASILLGNFSPDGYTLVEENGALKGKDCIL 653
Db 381 RHEPLSRWRDLRTIRAGISAVPLGSMALRQARMVLVGFSEGHVFEEDADCLTIGMHGRPL 440
OY 654 LTASAW 659
Db 441 FSASAW 446

```

RESULT 11

```

O9C8Y3 PRELIMINARY: PRT: 511 AA.
AC 09C8Y3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DE GIBBERELLIN REGULATORY PROTEIN, PUTATIVE.
GN T27F4.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kutz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Martelli A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
DR EMBL, AC020665; AAG52171.1;
SQ SEQUENCE 511 AA; 56754 MW; 1E60071697C92A9F CRC64;

```

Query Match 17.2%; Score 601; DB 10; Length 511;
 Best local similarity 34.6%; Pred. No. 1, 1e-31;
 Matches 146; Conservative 70; Mismatches 150; Indels 56; Gaps 10;

```

OY 273 QAAAAAAKERKEORRQKORE-----DESGHLTLTLLOCAEVNAD 313
Db 107 RAIPGSAVYPRDEHVTIRSKRTRIESELSTRSVYVLDSEGTGVRLVHALLACAAVOON 166
OY 314 NDDAHOITLEIETLPTGSTQORVAAVFAEAMSRALVSSCGLIYAPLPSPSPAARLH 373
Db 167 NLKLADALVYKHVGLLASOAGAMRKVATYFAGLARRI-----RIY 208

```

[illegible]

```

QY 433 SRGGPVPVRLTGTGASMEALEANGKRLSDPADPLGLGFEEFCANA -EKAGVWDEKIGV- 490
Db 427 LRPMPDPFRLTGTGISTLDIOWGWMKLGQASTIGVFEKRSIALNNLSLCKPEMLDIR 486
QY 491 TRREVAVAHMH--HSLVDVYGSDSNTLMLIORLAPKVYTWVEODLSHSGS-FLAREVEA 547
Db 487 PGLSEVAVNASVFEHLRLLAHFGSIDTKEFLSTKSTRPDLMTYVEQGANINGVFLDRFES 546
QY 548 IHVYSLEFSDLSADSYGEDSPERHVEQQLSREIRNVLAVGSPART-GDYKFGSKREKLA 606
Db 547 LHVYSLEFSDLEGPSOD----RVMSSELFRLQIINTLVACEGGEVERHERETLNWRNRFG 602
QY 607 QSGFRFAALASSAAQAASLLGMFP-SGCGYTLVEENGLKTKGKROLCLLTASAMR 660
Db 603 LGGFRPVISGNAXKQASMLLALYAGADGVNEENEGCLLIGWOTRLIATSAMR 657

RESULT 13
Q23642 PRELIMINARY; PRI: 587 AA.
AC 023642
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE RCAL PROTEIN.
GN RCAL
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97379310; PubMed=9237632;
RA Truong H.N., Caboche M., Daniel-Vedele F.;
RT "Sequence and characterization of two Arabidopsis thaliana cDNAs
RT isolated by functional complementation of a yeast gln3 gdh1 mutant.";
RU FEBS Lett. 410:213-218.(1997).
DR EMBL, Y11336; CAA72177.1;
DR Menzel, 24145; At4h13051;24145.
SQ SEQUENCE 587 AA: 64023 MW: DDA7A3C741FB51EF CRC64;

Query Match 17.0%; Score 594.5; DB 10; Length 587;
Best Local Similarity 31.8%; Pred. No.3.5e-31;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

QY 165 LTHNVEELIHPENPG-TASLELRLRSLAADPARLPDPPOQHALLHGAPAAAPGL- 222
Db 81 LSHLMTDVYH-YNPSELYSMLDNML-----SELNPPPLP-----ASSNGLD 120
QY 223 -TLPPP-----PPLPDKRRHHPPCQQQQQDEEPIRAPQSPKAPTAETAA- 268
Db 121 PVLPSEELIGFPASDYDLKVIPGMATYQFPALIDSSSSNNQKRLKSCSSPDSWVTSYST 180
QY 269 -----AAAQAQAAAAAAKERKEEQRKORDEGHLTLTLGLCAEVANADNLDDA 318
Db 181 GHOIGVGIVGTTTTTTTTTAAAESTRSYIIVDSQENGRLVHALMACAFALQONNLTLA 240
QY 319 HQTLEIAELATPFGTSYQVAVAYFAEAMASARLVSSGLGLVAPRPGSPAARLHGKRVAA 378
Db 241 EALVQIGIGLAVSQAAMKRVATYFEALARR-----IYRLSPQN---QIDHCTSD 289
QY 379 ARQV-FNGISPPVVKFSHTFANAQIQAFERREVRVHTLIDIMOGIOWPRLPHILASRPG 437
Db 290 TLQMEHYETCPYLKAHFANFANAQALLAEAFGKKRVHYIDFSMQGLOMPALMALRREG 349
QY 438 PPRVRLTGTGA-----SMEALFATGKRLSDFADTLGLPEFCA-VAEKANVDPKEKLV-- 490
Db 350 PPTFRLTGTGPPAPRPNDSHLHVEGCKLADLAELAIHVEFYRGFVANSIADLDASMLELR 409
QY 491 TRREVAVAHMH--HSLVDVYGSDSNTLMLIORLAPKVYTWVEODLSHSGS-FLAREVEA 547

```

```

Db      410 SPTFAVAVNSVFEHLKLLGRPGLEKVLGVYKQIKPVLFVTVEQESNHNQVFLDRFTES 469
      : |||||: : | | : : : : : | : |||: : | | : | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQQLSREIRNVLA VGGPART -GDVYFGSWREKLA 606
      : ||||| ||||| : : : : : : : : : : : : : : : :
Db      470 LHYSTLFDLSLE---GVPSQDKVMSVYLVGKQICNLVACGEPDVERHETLTSQMGNRFG 526
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      607 QSGFRAASLAGSAAQAASLLGMFPS -DGYTLVEENGALKGMKDLCLLTASAMR 660
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      527 SSGLAFAHLGSAVNAFKQASMLLSVFNSSGGYRVESNGCLMGMHTRPLITTSAMK 581
      : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 14
ID      023725      PRELIMINARY;      PRT;      587 AA.
AC      023725:
DT      01-JAN-1998 (Tremblrel. 05, Created)
DT      01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT      01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE      GRS PROTEIN.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=COLUMBIA;
RA      Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,
RA      Murphy G.P., Harberd N.P.;
RL      Genes Dev. 0:0-0(0).
DR      EMBL; Y15194; CAA75493.1; -
DR      Mendeil; 24071; Atach; 3051; 24071.
SQ      SEQUENCE 587 AA; 64006 MW; F6FGC738EE7DCA9 CRC64;

Query Match      16.9%; Score 593.5; DB 10; Length 587;
Best Local Similarity 31.8%; Pred. No. 4.1e-31;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

Qy      165 LIHNREIHPNPG-LASLELRSLAADPAPLPPOQOQHALLHGAPAAPAGL- 222
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      81 LSHLATDVH-YNPSELYSWLDNML-----SELNPPPLP-----ASSNGLD 120
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      223 -TLPP-----PLPDKRHNHPPPCQOQOQEEHPAPQSPKAPTAETAA- 268
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      121 PVLSPETICGFASPDYDLVIGMAIYQPAIDSSSSNNQKRLKSCSPDSMTSTST 180
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      269 -----AAAAQAAAAAAKKEKEORRKRQDEEGHLTLTLLOCAEAVNADNDDA 318
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      181 GTQIGVIGTIVTTTTTTTAAESTRSVLLVDSQENGVRVLAHLMACAIAIOONNLTLA 240
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      319 HQTLEIAELATPFGSTORVAAYFAEAMSAKLVSSCLGLVAPLPQSPAPARLHGVA 378
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      241 EALVQIGCLAVSQAGAMRKVATVFAEALAR-----YRLSPQN-----QIDHCLSD 289
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      379 AFQV-FNGISPFVKFSHTANQAOEAFEREERVHIIDIMQGLQWPGLEPHILASRPG 437
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      290 TLOMHFEYTCPLKFAHFTANQALIEAFEGKKRVHVIDFSMNOGLQWPAALQALALREG 349
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      438 PPRVRLTGLGA---SMEALATGKRLSDPADTLGLPFEFCA-VAEKAGVNDPEKLG- 490
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      350 PPTFRITGIGPPAPDNDSHLEHVGCKLAQLAELAVHEEYRGFAVNSLADIASMLRLP 409
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      491 TRREAVAVHMLH--HSLVDYTGSDSNTLMLQRLAPKVVTWVEODLSHSGS-FLARFVEA 547
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      410 SPTFAVAVNSVFEHLKLLGRPGLEKVLGVYKQIKPVLFVTVEQESNHNQVFLDRFTES 469
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQQLSREIRNVLA VGGPART -GDVYFGSWREKLA 606
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      470 LHYSTLFDLSLE---GVPSQDKVMSVYLVGKQICNLVACGEPDVERHETLTSQMGNRFG 526
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      607 QSGFRAASLAGSAAQAASLLGMFPS -DGYTLVEENGALKGMKDLCLLTASAMR 660
      : | | | | | : | | | | | : | | | | | : | | | | | :

```

```

Db      527 SSGLAFAHLGSAVNAFKQASMLLSVFNSSGGYRVESNGCLMGMHTRPLITTSAMK 581
      : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 15
ID      09SLH3      PRELIMINARY;      PRT;      587 AA.
AC      09SLH3:
DT      01-MAY-2000 (Tremblrel. 13, Created)
DT      01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT      01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE      PUTATIVE GCAL, GIBBERELLIN RESPONSE MODULATION PROTEIN.
GN      AT2G01570.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA;
RA      MEDLINE-20083487; PubMed-10617197;
RA      Lin X., Kaul S., Rounsley S.D., Shea T.P., Beilto M.-I., Town C.D.,
RA      Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA      Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA      Cronin L.A., Shen M., Vanaken S.F., Umayam L., Tallon L.J., Gill J.E.,
RA      Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA      Coppenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA      Salzberg S.L., Fraser C.M., Venter J.C.;
RT      "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL      Nature 402:761-768(1999).
DR      EMBL; AC005560; AAC67333.1; -
SQ      SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;

Query Match      16.9%; Score 593.5; DB 10; Length 587;
Best Local Similarity 31.8%; Pred. No. 4.1e-31;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

Qy      165 LIHNREIHPNPG-LASLELRSLAADPAPLPPOQOQHALLHGAPAAPAGL- 222
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      81 LSHLATDVH-YNPSELYSWLDNML-----SELNPPPLP-----ASSNGLD 120
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      223 -TLPP-----PLPDKRHNHPPPCQOQOQEEHPAPQSPKAPTAETAA- 268
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      121 PVLSPETICGFASPDYDLVIGMAIYQPAIDSSSSNNQKRLKSCSPDSMTSTST 180
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      269 -----AAAAQAAAAAAKKEKEORRKRQDEEGHLTLTLLOCAEAVNADNDDA 318
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      181 GTQIGVIGTIVTTTTTTTAAESTRSVLLVDSQENGVRVLAHLMACAIAIOONNLTLA 240
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      319 HQTLEIAELATPFGSTORVAAYFAEAMSAKLVSSCLGLVAPLPQSPAPARLHGVA 378
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      241 EALVQIGCLAVSQAGAMRKVATVFAEALAR-----YRLSPQN-----QIDHCLSD 289
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      379 AFQV-FNGISPFVKFSHTANQAOEAFEREERVHIIDIMQGLQWPGLEPHILASRPG 437
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      290 TLOMHFEYTCPLKFAHFTANQALIEAFEGKKRVHVIDFSMNOGLQWPAALQALALREG 349
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      438 PPRVRLTGLGA---SMEALATGKRLSDPADTLGLPFEFCA-VAEKAGVNDPEKLG- 490
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      350 PPTFRITGIGPPAPDNDSHLEHVGCKLAQLAELAVHEEYRGFAVNSLADIASMLRLP 409
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      491 TRREAVAVHMLH--HSLVDYTGSDSNTLMLQRLAPKVVTWVEODLSHSGS-FLARFVEA 547
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      410 SPTFAVAVNSVFEHLKLLGRPGLEKVLGVYKQIKPVLFVTVEQESNHNQVFLDRFTES 469
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQQLSREIRNVLA VGGPART -GDVYFGSWREKLA 606
      : | | | | | : | | | | | : | | | | | : | | | | | :
Db      470 LHYSTLFDLSLE---GVPSQDKVMSVYLVGKQICNLVACGEPDVERHETLTSQMGNRFG 526
      : | | | | | : | | | | | : | | | | | : | | | | | :
Qy      607 QSGFRAASLAGSAAQAASLLGMFPS -DGYTLVEENGALKGMKDLCLLTASAMR 660
      : | | | | | : | | | | | : | | | | | : | | | | | :

```

Tue Apr 23 11:12:24 2002

Db 527 SSGIAPAHIGSNAPKQASMLSVFNSGQGYRVEESNGCLMLGWHTRPLITTSAMK 581

Search completed: April 22, 2002, 19:42:46
Job time: 571 sec

REFERENCE	2 (bases 1 to 3100)
AUTHORS	Llm,J., Helarnta,Y., Specht,C.D., Jung,J., Sims,L., Bruce,W.B., Diehn,J., and Benfey,P.N.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAY-2000) Biology, New York University, 100 Washington Square East, New York, NY 10003, USA
FEATURES	location/Qualifiers
source	1..3100

```

organism="zea mays"
db_xref="taxon:4577"
join1<1..1562,2410..>(2854)
/gene="SCR"
/product="SCARPCROW"
<1..>(2854)
/gene="SCR"
join1<1..1562,2410..>(2854)
CDS
gene

```

```

/gene="SCR"
/function="transcription factor"
/note="rzmSCR"
/codon_start=1
/product="SCARECROW"
/protein_id="ANG13663.1"
/db_xref="GI:10178637"

```

/translaction="MPPPPPPPLTTTCRCRCPPPHLPPEPPSSNHHLLHYLHLQHQ
 EAAHAAVRRKRASPDMDLPPPRHVVGDSLVNVAHAAAGGSGASPSASAALPALT
 QHLOPLAFQHNHAEVYVPAHPAPAAHQAQSGERTASTTAMVCGITFDITLGGSGAV
 STIOLHNVREITLHPCCNGLASLELLEKRLADPAPPPPPPOQHALLHQAHPAA
 PAGITLHPPEPLDPKRRHNEHPPCQOQOQDEHPADPSAPPAETEAHAAHAAQAA
 AAAKERKEEQRKQRQDEBGLHLLTLLQCAEVANADNLDDAQTLEIKELTPEST
 STQVAVAFAMASQARVSSCLGJLPLPGSPAAALHGVAAEEVGNISFPEV
 SHFANAOIODEFERERHVIIDJIDMOGQOMPELFIILASRPGCPVRLTGTGAGM
 EALEATGKRLIDFADDTGLPPEFCVAKENGNDPELTGVTGRVAPVHLTHSLYV
 TGSQNTLITLQRLAKVYVMWEDJSHSGSLAPFEALHYHSALEDSIDSAZGDS
 PERHYVQOQLSREIRNVLAVGGPARGTQVFGSGREKTLAQSGFRASLSLGSAAQAS
 LLLGMPSPDSGTLLEENGALTKLQKDKDCLLLTASAKWRITQVPPCR"

Query Match	88.3%	Score 3100;	DB 8;	Length 3100;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 3100; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	293	atgcacacgcacacgccttcgccttccttcacttccttatatgcccgcgcgtgcctctcccca	352
Db	1	ATGGCACCAGCACCAGGCTCCGGCTCTCCTCTCACTCTTATTGGCGCGCTCCCTCCCCA	60
QY	353	caactcccttcgcgccttccttccttcttccccaacaacttccttccttcacttccttcag	412
Db	61	CACCTCCCTCGGCTCCTCTCTTCTTCCCAACAACACTTCTCTCCACTACTCATCAG	120
QY	413	ctagaaccaacaagaagccgcgcgcgcgcacatggttcgcaagcgcgcgcgcgtccgata	472
Db	121	CTAAGCACCAAGAAAGCGCGCGCGCCATGGTTCGCAAAGGCGCCCGGTCGCACATG	180
QY	473	gaactccgcgcgcgcgcgcacatgtaagggcgaactctcgcagcttcacgcgcgcgcct	532
Db	181	GACCTCCGCGCGCGCGCGCGCAGCTACGCGGACCTCTCCGACGTCAAGCGCGCGCT	240
QY	533	gcgcgcgcgttgttggtagtggcgcgcgttccttcgcgcgcgcgcgcgcgcgcgcgt	592
Db	241	GCGCGCGGTGTTGGTGGTAGTGGCGCGCGCTCTCCGACGCGGCGAGCTGCCCGCGCTG	300
QY	593	cccaaccagctccacacgcgcgtgcctcccggttcagagcaacgcgcgcgcgcgcgcgc	652
Db	301	CCGACCCCAAGCTCCACAGCTGCCCCCCGCTTCCAGCACACCGCGCGGAGTGGAGCTG	360
QY	653	ccgcgcacacccgc	712
Db	361	CCCGCGACACCGCGCGCGCGCGCCACAGCGCAGCGCGCGCGCGCGGCGGAGGACACCGCTCCAGC	420
QY	713	accgcgtgggtggagcgcgcgcacatcctccgcgcacatcctcggaacgcgcgcgcgcgcgc	772

Db	421	ACCGCGGGGAGACGGGCATCATTCGCGACATCATTCGGAGGACGAGGGGGCGCGGATC	480
Oy	773	tcacatcagcagctcatcacaacgctccgagatcatcacccctgcaaccggcctc	832
Db	481	TCCATTCACGGACAGCTCATCCCAACAGTCCCGAGATCATCATCCCTCGAACCCGGGCTC	540
Oy	833	ggctcagctccctggagctccgcctcgcctccctccctcagccgagcccgcccaactgcg	892
Db	541	GGGTGCGTCTCTGGAGCTCCGGCTTCGGCTTCCTCTCTCGACAGCCGACCCGGCCCACTGGCG	600
Oy	893	ccgcgcgcgcgcagccgcagcaagcaltgctctctctgcaagcgctccggccgcctccgcg	952
Db	601	CCGGCGCGCGGACGCGGAGCGCATGCTCTCTTGCAACGGGGCTTCGGCGCGCTCCGGCG	660
Oy	953	ggagttacagctctcctccccccgcacacgcgtctcggagcaaacgcgcgcacagagcatcacgcg	1012
Db	661	GGGCTGACGCTCTCCTCCCGCCGACCGCTTTCGGACACAGCGCCGCCACAGCATTCACCG	720
Oy	1013	ccgctgcacagcagcaacaacacagagagagacccgcaltcggcgccgcagtcgcccagaagcccg	1072
Db	721	CCGTGCGCAGCAGCAACAGCAGAGAGAAACCGCATCCGGCGCCGACATCCCAAGGCCCG	780
Oy	1073	accgcggaagaagacacgcagcgcgcgccgcgcgcgcgcgcacaaagcagcagctgcgtcgccgc	1132
Db	781	ACCCGGGAAGAGACCCGACGCGCGCGCGCGCGCGCGCAACAGCAGCAGCTGCTCGCGCGCC	840
Oy	1133	aagcagcggagagagagagcagcgcgcggaagcagcgcgagagagagagcctccactgctg	1192
Db	841	AAGGAGGGGAAGAGAGAGAGCAGCGCGGAAAGCAGCGGACAGAGAGAGGCGCTCCACTGCTG	900
Oy	1193	acgcctgctgctgctcagtgctgcgcgcgcgcgtgaaacgcgcgcgcgcgcgcgcgcgcgcgcgc	1252
Db	901	ACCGTGTGCTGCTGCAATGTCGCCCGAGGCCGTGTAAACGGCGAACCTTGAGACGCGACACAG	960
Oy	1253	acgcctgctcggagatcgc	1312
Db	961	ACGCTGCTGGAGATCCGGGAGCTAAGGAGACGCCGTTTCGCACTTGACGACGACGCGGTGGCC	1020
Oy	1313	gctcactctcgc	1372
Db	1021	GCTCACTTCGCGGAGGCGCATGTGCGCGCGCTCTGTCAAGTCTCTGCGCTTGACACCG	1080
Oy	1373	ccgcctgc	1432
Db	1081	CCGCTGCGCGCGGGGCTCCCGCGCGCGCGCGCGCTTCACAGCGCGCGCTGCGCGCGCTTC	1140
Oy	1433	cagggtgtlcaacaggcatcagccccccttcgtlcaagtltctgcacatlcaaccgcacaacaggcc	1492
Db	1141	CAGGTGTTCAACGGCATCACGCCCTTCGTCAGATTTCTGCATTCACGCCCAACAGAGCC	1200
Oy	1493	atcccaaggagcgcttcgagcgcgagagagcgcgctgcaacatcatctgcaactcgacatcatgag	1552
Db	1201	ATCCAGAGAGCGGTCGAGCGGGAGGAGCGCGTGCATATCATGACCTTCGACATCATGAG	1260
Oy	1553	ggagctgcagtgccgggggcgttcctcaacatcctgtgccctccgcgcgcgcgcgcgcgcgcgcgcgc	1612
Db	1261	GGGCTGACAGTGGCGGGGCTTTCACATCTCTTGCTTCCGGCCCGGGGGGCGCGCCAGG	1320
Oy	1613	gttagagctcacccgcgcctcggggcgctccatlgtagagcgctcagagccaaaggagcctc	1672
Db	1321	GTGAGGCTCACCGGCTCTCGGGGCTCATAGAGAGCGCTTCGAAGCCACAGGGGAAGCGCTTC	1380
Oy	1673	tcgcatttcgcagacacgc	1732
Db	1381	TCCGATTTCCGCGACAGCTCGGCGCTTCGCCCTTGGAATTTGCGCCGTGCGCGAAGGCC	1440
Oy	1733	ggcaaatgtltaaccggagaaagctcagggtgtlcaagagcgagagcgagagcgctgcgcgtccatctg	1792
Db	1441	GGCAATGTTGACCGGAGAGCTAAGGGGTACAGAGGCGGAGGCGGTGCGCTCCACATGAG	1500
Oy	1793	ctgcacacactcgctctlaagcagctcactgctccgcactccacaagcgtctgctcatccaa	1852
Db	1501	CTGACACACTGCTCTTACGACGCTCACTGGCTCCGATTCACACGCTCTGGCTCATCCAA	1560

QY 1853 aggttagaaggaagtaacacatctctcgatccctgactctctgtctacatgtlcaaatcttg 1912
 |||||
 Db 1561 AGGTAGGAGAGATACACATCTCTCGATCCTGTACTCTTGTGTACCATGTCAATCTTG 1620
 QY 1913 atgcaatcagcgcaacttttaagctactaaacttggtttagcgaatgcgaatcagat 1972
 |||||
 Db 1631 ATGCAATCATGCGCACTTTCAGTACTTAACACTTGTAGTTCAGCAATGCGACATCCAGT 1680
 QY 1973 acaactaatcctaaaaaataactctcagaaggttccctaagtaaaaaacgcggttttgga 2032
 |||||
 Db 1681 ACAACTATCTAAAAAATATCTTCAGAGGTTCTCAGTAAAAAACCCGCTTTTGGGA 1740
 QY 2033 gctcaaaaagcttgatcattatgacaaacaaacttcttaaggcttaaaaaaggttgaaatctg 2092
 |||||
 Db 1741 GCTAAAAAGCTTCTCATTTATGACCAACCACTTTCAGGCTTAAAAAGGTGAATCTTG 1800
 QY 2093 gcaatgcttttgagaagatgctgtacgaagatctgtgtaagagagatctctcattgccc 2152
 |||||
 Db 1801 GCAATGCTTTTGACACGATGCTGTACTGAACTGTGAGAGAGATATCCCTCCATGCGC 1860
 QY 2153 ttgttgatcccaagaacacaaaaagatagatatttcgctcgcatltgtgtagtgaggtgt 2212
 |||||
 Db 1861 TTGTGTATCCAGACACACAAAAAGATAGTATTTCGCTCGATTTGGTTAGTGAAGGTGT 1920
 QY 2213 tctgatacacttgagagatgagctggaaggttccatcatcatgacgaacttccatg 2272
 |||||
 Db 1921 TCTATATCATCTTGAGAGATGAGAGTGAAGTCTTATCATCATGACCAACTTTCATG 1980
 QY 2273 gcaaaaaggttcttagtccaaagtgcagaagcgatgattctgaaagctgaaagaaataa 2332
 |||||
 Db 1981 GCAAAAGGTTTCTGATTCGCAAGTGGCAGAGCATGATTACTGAGTGCCTGATGAGAACTAA 2040
 QY 2333 ctgtcatctcttaccactaaacatcattatlaataacaaatcatcatccgagacctaa 2392
 |||||
 Db 2041 CTGTCACTCTTCTACCACTAACCATCATTTATTATACATAATCATCATCGGAGCCTAA 2100
 QY 2393 actaagaagaagctatacaaaagtgcattcttccaaatgctgcgcatatgccaagtgtga 2452
 |||||
 Db 2101 ACTAGAAAGGCTAATCAAAAGTGCAATCTTTCCAATGCGCCATATGACCAAGTGTGA 2160
 QY 2453 catgctggccatgtaacttcttgcgtgaaaccatctgcgtctcaagacatgagaaagcc 2512
 |||||
 Db 2161 CATGCGTGGCATGTACTTTTGGTGGAACATCTGCTCAAGCATGAGATGAAAGGCC 2220
 QY 2513 tgaactgcgaatgctctgatttgatgcaaccattatagaagaacgctlaagcgatgccc 2572
 |||||
 Db 2221 TGAAGTCAATGTCTTGATTTGATGCAACCATATTATGAAAGAACGCTAAGGATGCCG 2280
 QY 2573 gtccctggcaagggaatgcatatcgtaacacgaacgaaggaatccggaatcgaaatgctag 2632
 |||||
 Db 2281 GTCTGGCAAGGCAATGCAATATGTCAGACAGACAGGGAATTCGGGAATGGGCTAG 2340
 QY 2633 ctgtgtgcaaatgcgaaggaatatacaactacatggtgcatgatttccatccccaca 2692
 |||||
 Db 2341 CTGGTGAACAATGCGACAGGGGATTAATAAACTACATTGGTCATTGATTCATCCCCACA 2400
 QY 2693 caectgcaagctggcccccaagtgtgtgacaatggtggaacgaagacctggaacctcgag 2752
 |||||
 Db 2401 CACTGCAAGGCTGGCCCCCAAGGTGTGACAAATGTGAGACAGAGACCTGAGCACTGGGG 2460
 QY 2753 ctctctcctggcgagctgtgtgagagccatcactactactgagcgctgtgtgaactcgct 2812
 |||||
 Db 2461 CTGCTTCTTGCGCGCTTCGTGGAGGCAATCCACTACTCTGCGCTGTGCTGACTGCT 2520
 QY 2813 ggaagcagagctagcggaaggaacagccccagcgcaactgctgtgaggaagcagctgtgtc 2872
 |||||
 Db 2521 GGAAGCGAGCTAGCGGAGGAGCAAGCCCGAGCGGCAAGCTGTGAGAGAGAGAGCTGCTGTC 2580
 QY 2873 gcgggaaatccgcaagctgctgtgcgtgtggcgagcgccgcaacggcggaagctcaagtt 2932
 |||||
 Db 2581 GCGGGAATCCGCAACGTTGCTGGCGGTGGCGGGCGGCCCGCACCGGCGACGCTCAAGTT 2640

QY 2933 cgcgcagctggcgcggaagaagctggcgagctccgggttccgcgcgcgcctgcctgcgcgcag 2992
 |||||
 Db 2641 CGGCACCTGGCGCGGAGAAAGCTGGCGGAGTTCGGGTTCCGCGCGCGCTGCTGCGCGGAG 2700
 QY 2993 cgcgcgcgcgcgcgcgcgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgc 3052
 |||||
 Db 2701 CGCGCGGCGGCGGAGCGCTCCCTGCTGCTGCGCATGTGTTCCCTCCGACGGGTACACCTGT 2760
 QY 3053 ggaaggaagc 3112
 |||||
 Db 2761 GGAAGGAGAACGGCGCGCTTACAGCTCGGCTGGAAGAGACCTCTGCTGCTACCGCTGCGC 2820
 QY 3113 ctggcgcccatccaaagtgccgcgcgctgtgataagaaacctctgcctgcctgcctgc 3172
 |||||
 Db 2821 CTGCGCCCGCCATCCAGAGTCCCGCGCGCTGATGAGACCTCTGCTGCTGCTGCTGCTG 2880
 QY 3173 gttgagaagcgccgc 3232
 |||||
 Db 2881 GTTGAAGGCGCGCATCCACTGTTGTTTCATCTGTAGCTGCTGCTGCTGCTGCTGCTGCT 2940
 QY 3233 ggaagataagaaagcggaacgtaactaattgctcctgagatagatccatccatcagat 3292
 |||||
 Db 2941 GGGAGATTAAGAAAGCGGAAAGCGTACTATGCTCTGAGATGATCCATTCATTCAGT 3000
 QY 3293 gatagttactgatactaaagcttaattagttcaatgtagatgctgctgttcaagtg 3352
 |||||
 Db 3001 GATAGTTACTGATGTACTAAGCTTTAATTAATGTTCAATGCTAGATGCTTCTTCAAGTG 3060
 QY 3353 tgcagtcgatacctgtcctgtcctgtccttcttcttcttcttcttcttcttcttcttct 3392
 |||||
 Db 3061 TCGATGCGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3100

RESULT 2
 ATE24B22
 LOCUS ATE24B22 100285 bp DNA PLN 28-JAN-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F24B22.
 ACCESSION AL132957
 VERSION AL132957.1 GI:6434216
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 100285)
 REFERENCE
 AUTHORS Bloeker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetlier, F. and Salanoubat, M.
 Unpublished
 JOURNAL
 REFERENCE
 2 (bases 1 to 100285)
 AUTHORS EU Arabidopsis sequencing project.
 TITLES Direct Submission
 JOURNAL Submitted (28-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, project-coordinator: Marcel Salanoubat and Francis Quetlier, Gropement d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr

COMMENT
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

FEATURES
 source
 1..100285
 Location/Qualifiers

/organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"

misc-feature
 1..524
 /note="Various EST matches (T44303, H36397, N97294) without corresponding genemodel"
 2053..3629
 /gene="F24B22.10"

CDS
join(2053. .2499,2625. .2852,2948. .3454,3558. .3629)
/gene="F24B22.10"
/note="Contains Fructose-1-6-bisphosphatase active site
AA358-370; Prokaryotic membrane lipoprotein lipid
attachment site AA101-111"
/codon_start=1
/product="fructose-bisphosphatase precursor"
/protein_id="CAB70979.1"
/db_xref="GI:6822051"
/translation="MAATTAATTTSSHLLSSSRHVASSSQPTLSLSPRSLSNNKKRAP
TGVNRHGVASGVRCAVAADAETKTARKSGELETGLGMLLRQEMGEIDAELETL
VMSISILACKQIALYORAGISLTVGVGVANVIOGEDOKLIDVINSNEVSNCRSSGR
TGLTASEEDVPVAVEESYSGNVYVDPDGSSSIDAAVSTGIFGYSFETDPMG
DSDDISALGSEORCIYVNCOPGNMLLACYCWSVIEFLTLGKVFETDLVPG
EFVLTQENITPKRGRIVSPNEGVMMDKIKYIDDLKDPGPKVPSARYIGSLV
GDFHRTLLYGGIYICPPDAKSKNKKLLLECAPMSTIVDAQGKSDGHSRVLDP
TEIHQRPVLTIGSTEVEKLEKTLA"
/number=1
/gene="F24B22.10"
/number=1
2500. .2624
/gene="F24B22.10"
/number=1
2625. .2852
/gene="F24B22.10"
/number=2
2853. .2947
/gene="F24B22.10"
/number=2
2948. .3454
/gene="F24B22.10"
/number=3
3455. .3557
/gene="F24B22.10"
/number=3
3558. .3629
/gene="F24B22.10"
/number=4
4017. .5928
/gene="F24B22.20"
complement(4017. .4082)
/number=1
complement(join(4017. .4082,4178. .4262,4350. .4977,
5060. .5242,5424. .5658,5755. .5928))
/note="similarity to mucin MUC3, Homo sapiens,
EMBL:AF007193"
/codon_start=1
/product="putative protein"
/protein_id="CAB70980.1"
/db_xref="GI:6822052"
/translation="MGKSSRSKSGNLIGKEVPTQVAFIVDRYLHNDRSETRALFR
SEASLISDPIRMVPLMTLDAMLNHYSLKOKYSLDOEKLKLDQEKIRQNLQ
GEMVMTYNASLTAPPASAPTSQOKNHSISSGSLQYNTLNGMSVILKGNRVG
NESTPSPSOSITGCRKGRPEVSATPSPSRKSRITRANGTKLPOADKAANPTSETLA
VAKASANEILGNSSVYKCLFNKADSSVPTSTCEFTPOKHAASSGSDKSNSSQKEVT
PNTNCTIVTERKPTISPLKQITSYVERSHLLSFSSPVKSNLKMKNKRVHVGKLFN
DDTDTETCLEAPADADVSTSPSGSEVEDLFPMDFENLDFSELVDFDGCSSANH
CISLTNPQPTQTVGSSSPESGDCNLESDDQFLFETSTVTDVIOISRCIDPLSNSSHS
LNTKXI"
complement(4083. .4177)
/number=1
complement(4178. .4262)
/number=2
complement(4263. .4349)
/number=2
complement(4350. .4977)
/number=3
complement(4978. .5059)
/number=3
complement(5060. .5242)
/number=4
complement(5243. .5423)
/number=4

exon
complement(5424. .5658)
/number=5
complement(5659. .5754)
/number=5
complement(5755. .5928)
/number=6
6432. .6705
/gene="F24B22.30"
complement(join(6432. .7073,7206. .7519,7744. .7911,
8003. .8470,8573. .8705))
/note="similarity to delta-latroinsectotoxin precursor,
L.mactans, EMBL:LMTDLATOX"
/codon_start=1
/product="putative protein"
/protein_id="CAB70981.1"
/db_xref="GI:6822053"
/translation="MGITDKGQQLQOEYIVYEVGPPSSSPNEPPSSLPNVEDTSED
VTKPSRLNLTAKVLTGMKTASTLISKECNVBEQITGNSDLALHIAVAKHKDFVR
NLRLREMDPDLISLKNKDGNTPLSFPAALGDIETAEMLINMLRDLPDINSEKTMPIH
IALYGHGEWQYLFPSKTSIKDLNDQOYLNFHTMSADIVGVADVPLMLERVDLYR
KELALYPSNKRKALHLARTSAISHRSQNLFOQVASSMLFDDAELGNVILYLIR
SHLDLIMVYDNNRNLTFVVAALYRHEINLESLIYELGKKDLIASYKEKQSDTLILY
ARLPNNRQVSGGAAIHLQKELILFKAVKEIVPSYIETNTKCELANDHIFTEQHEN
LRRKEGRWKETATACMLGATLIATVRAAATTPGGNDSDGKANTLGFENRRL
FDIFLSDSVALFSSMSIVIFLSIFTSRYAEEDRYDLPKLMFGLSALFISISM
LAFTESMILIRVEKASLSLVLSCLASLTALTFAVLYFHLMFNLRVYISMFLGR
KRGIRLEFV"
complement(6432. .7073)
/number=1
complement(7074. .7205)
/number=1
complement(7206. .7519)
/number=2
complement(7520. .7743)
/number=2
complement(7744. .7911)
/number=3
complement(7912. .8002)
/number=3
complement(8003. .8470)
/number=4
complement(8471. .8572)
/number=4
complement(8573. .8705)
/number=5
10499. .11551
/gene="F24B22.40"
10499. .11551
/gene="F24B22.40"
/note="similarity to probable lectin 2 precursor, alfalfa,
PIR:T09620"
/codon_start=1
/product="putative protein"
/protein_id="CAB70982.1"
/db_xref="GI:6822054"
/translation="WASFLISGLLVLLSLTALLSISISISATSSSFSSGKNASR
EESDIALYGDALVDGGSILQITDVSHGGRVLYKPIESNMNNTKEVYFGFTFS
FSISPSRGGRLEFVFPVNETFDHSLFQVKEDTSPFOIGDSNVAVVDATVSERI
RNFTITANLEKTEKVLIVAMINVOAGKFELEVLSKSFESVLLPLMFQDIDLSOMLR
EDEPVAVNSYSGNVNLHMSLSLEVRHSYEBSMAPVLEBOLRBEOLRKEAARRT
SPRMREVYTCFYMTGSGTGLVFFAMHIIYAAFKRNRLAMVNOECGIKTKRFGYKKME
KKEVYTSNADAKQERR"
/number=1
13247. .13969
/gene="F24B22.40"
/number=1
13247. .13969
/gene="F24B22.50"
/number=1
13247. .14937
/gene="F24B22.50"
join(13247. .13969,14245. .14937)
/gene="F24B22.50"
/note="similarity to fructokinase, Lycopersicon


```

/protein_id="AAB06318.1"
/db_xref="GI:1497987"
/translation="MAESGDENGGOPPPHSPRLRTSSSSSSNNKRGPPPPPPPLVNV
RKRLASEMSSNPDYNNSSRRPRRVSHLDSNTVNTVQPPSLTAATVSSQPPPLIS
VGFSGFLVFPSSDRGNWMSVQPMODSSSSSSAPTWADAIIRLIISSVSIP
QIIONVRDIIPCPNMGALIELRLISLMLDPSSSDPSPFEPPELYQISNNPPO
QOOHQOQOQOQHPPPPQQOERENSTDEAPPELVTPVONTVALEERKE
EIKROKODEEHLHLTLLOCAEAVSADNLEAKMLLETISOLSTPYGTSARORVAAE
SEAMASRLNSCLGITAALPRMPPOTHSIKMVARQVNGISPLVFSFPTNOAO
EAFKEDSVHIIIDIDIMOGLOMPCFIHLISRECPPHVRLIGLISMEALQNGKRL
SDFTDKLGPPEFCPLAEKGNLDTERLVNKRRAVAHMLQHSILVDYSDHATML
LQRLAPRVAVVEQDLSHAGSFLGRFEVALHYVSAIPDSLGASGESEERHVEQD
LSKEIRNVLAIVGSPRSRGSEVYFESMRKMOOCGRGISLGNANATQITLLGMFSPDQ
YTLVDNDNTEKLGKMDLSLLTASAMTPRS"
intron
1581..1695
/gene="SCARECROW1"
/exon
1696..2125
/gene="SCARECROW1"
BASE COUNT 571 a 546 c 439 g 607 t
ORIGIN
Query Match 10.2% Score 358.4; DB 8; Length 2163;
Best Local Similarity 39.38; Pred. No. 1e-33;
Matches 696; Conservative 0; Mismatches 451; Indels 27; Gaps 4;
QY 702 ccgcgtccacagccgcgtggtgagcgcacatccgcacatcatcgagcagcg 761
DB 455 CTGCTTCACTACCTATAGGTGTGACGCCATTATCAGACCTTAT-----CCATTCT 508
QY 762 gcgcgcgggtctcacaacagcgtcacaacgctcgcgagatcatccaccctgca 821
DB 509 CAACTTCAGTCTCTATCTCAACTTATCAAAAGCTTAGAGACATATCTTCCCTGTGA 568
QY 822 accccggctcgcgcgtcctctgagctccgcctccctcctcgcgcgcgcgcg 881
DB 569 ACCCAAAATCTGGTCTCTTCTTGAATACAGGCTCCGATCTCATGCTCTTATCTT 628
QY 882 ccccaatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 941
DB 629 CCTTCTCTGACCTTCTCTCTCAACTTTCGAACCTCTTATCAGATCTCCCAACATC 688
QY 942 ccgcgtccgcggggcgtgagctccctcccccgcacgcgtcgcgacagcgccgcag 1001
DB 689 CTTCCTCT-----CCACAAACAGCAACACACACCAACAAACAAACAAAC 733
QY 1002 agcatccacccgcgtgacagcagcagcagcagcagcagcagcagcagcagcag 1061
DB 734 AGCATTAAGCTCTCTCTCTCCGATTTCAGCAGCAAGAAAGAAATTTCTTACCGATG 793
QY 1062 ccaagcccccagccgcggaagagacccgcgcgcgcgcgcgcgcgcgcgcgcagctg 1121
DB 794 CACCAACCCCAACAGACAGACAGTGAACGGCCACTGTTCCCGCGCTC---CAACAATATCG 850
QY 1122 ctgcgcgcgcgaagagcgaagagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1181
DB 851 CGAGAGCTTTAAGAGAGAGAGAAAGATTAAAGAGCAGCAACAAACAGACCAAGAGAT 910
QY 1182 tccacctgtagcgtgtgctgtagcagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1241
DB 911 TACACCTTCTCATATGCTGCTACAGTGTGCTGAAGCTGTCTCTGCTGATATATCTCAAG 970
QY 1242 acgcgcacacagcagcgtgtagagatcgcgagcagcagcgcgttcgcacctcgcagc 1301
DB 971 AAGCAAAACAGCTTCTTCTTGAATCTCTCAGTTATCATCACTCTTACGGGACCTCACGC 1030
QY 1302 agcgcgtagcgcgtactctgtagcagcagcagcagcagcagcagcagcagcagcagc 1361
DB 1031 AGAGAGTAGAGTGTACTTCTCGAAGACTATGAGCGAGATTACTCAACTCGTGTCTCG 1090
QY 1362 gccgtatagcgcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1421
DB 1091 GAATTTACGGCGCTTGTGCTTACGCGTGATGCTCAAAACGCAATAGCTTGAATAATGCTCT 1150

```

```

QY 1422 ccgcgcgtctccagcagcgtgttcaacgcacagcccccctgcagcgtctcgcacattcacg 1481
DB 1151 CT---GGCTTCAAGGTCTTTTAATGGATTAAGCCCTTAAAGAAATTCACATCTTAAAG 1207
QY 1482 ccaacgcgcacatcccaagcagcgtctcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1541
DB 1208 CGAATCAGCGGATTCAGAGCATTTGAGAAAGACAGACTGTACATCATTCATTGACTTGG 1267
QY 1542 acatctgacggcgtcagtcagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1601
DB 1268 ACATCATGACGGGACCTTCATGCTGCTTATTCACATTCCTTCTTACACCTGAG 1327
QY 1602 gcccccgcagcagcgtgagcgtcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1661
DB 1328 GACCTTCACACGTGCGACTCAGCGACTGGTGTGATCTTCATGAGAGCTTTCAGGCTACAG 1387
QY 1662 ggaagcgcctcctcgcatttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1721
DB 1388 GGAAGCGTCTTTCGATTCACAGATTAAGCTTGCCCTTGCTTGTGATTCGCTTGA 1447
QY 1722 ccgagaagcgcgcgaatgttgaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1781
DB 1448 CTGAGAAAGTTGAAACTTGACACCTGAGAGACTCAATGTGAGAGAAAGGAAAGCTGTGG 1507
QY 1782 ccgtcacctgctgcacacactgcctctacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1841
DB 1508 CTGTTCACCTGCTTCACATCTCTTTATGATGTCTCACTGGCTCTGATGCACACACTCTCT 1567
QY 1842 ggcctcacaagagtagaagagtagacacatct 1875
DB 1568 GGTACTCTCAAGTAAATTAACATTAACCTTT 1601

```

```

RESULT 4
AP003866 LOCUS
AP003866 137775 bp DNA HTG 10-JUL-2001
DEFINITION
Oryza sativa chromosome 7 clone OJ1092_A07, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AP003866
VERSION
AP003866.1 GI:14646799
KEYWORDS
HTG; HTGS-PHASE2.
SOURCE
Oryza sativa (cultivar: Nipponbare) DNA, clone: OJ1092_A07.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 137775)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone: OJ1092_A07
Published only in database (2001) In press
2 (bases 1 to 137775)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submision
Submitted (09-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@abrr.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
FEATURES

```

```
source
1. 137775
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="01092.A07"
BASE COUNT 38997 a 30139 c 29847 g 38726 t 66 others
ORIGIN

Query Match 10.0%; Score 350.2; DB 2; Length 137775;
Best Local Similarity 70.5%; Pred. No. 1.6e-33;
Matches 514; Conservative 0; Mismatches 203; Indels 12; Gaps 3;

OY 1135 ggaagcgaagaagagagcagcgagcgaagcagcgaagagagcctccacactgctgac 1194
Db 74386 GGTGAGTACGAGGAGGAGGAGGAGCGGCGGAGACCGCGGCTGGCTGGCTGCTTAG 74445
OY 1195 gctgctgctgagtcgagcgagcgctgagacgagacacactggaagcagcagcagac 1254
Db 74446 TTTGTTCTGAGGTGTCGGAGGCGGTGGCGATGACCACTGCCGAGGCGGGGACT 74505
OY 1255 gctgctgagatcgcgagcagcagcagcagcagcagcagcagcagcagcagcagc 1314
Db 74506 GCTGCCGGAATCCCGACGCTGGCGTCCGCTTCCGCTCCGCGCGCGCTCGCGGC 74565
OY 1315 ctacttcgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1374
Db 74566 CTACTTCGGGGACGCGCTGCGCGCGCGCTGAGAGCTCTACTCGGGGCTCACTCGCC 74625
OY 1375 gctgagcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1434
Db 74626 GCTGGCGCTCCGCGCGCTGCGCGCGCGCGAGAGC--GCGGATCTCCGCGCGTTCCA 74682
OY 1435 ggtgttcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1494
Db 74683 GGCCTACAGCGCGCTGTCGCGCGCTGTCAGATGCTTCGCACTTCAGCGCAACAGGCA 74742
OY 1495 cgaagagcgcttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1554
Db 74743 CTTCACAGCGCTGTCAGCGAGCGCGCTGTCAGATGCTTCGCACTTCAGCGAGG 74802
OY 1555 gctgagcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1614
Db 74803 GCTCAAGTGGCGCGCTCTTCCACATCTCGCTCCCGCCCAACAGCGCGCTCGCT 74862
OY 1615 gaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1674
Db 74863 CCGGATCACGGCGCTCGCGCGCTGTCGAGCTGCTCGAGGCCACCGCGCGCTCGC 74922
OY 1675 cgatttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1734
Db 74923 CGACTTCGCGCGCTGCTCGCGCTGCTCGAGTTCCGCGCCATCGAGGGGAATCGG 74982
OY 1735 caatgt--tgaccggaagaagcagcagcagcagcagcagcagcagcagcagcagcagc 1785
Db 74983 GCACGTGCCGAGCGCGCGCGCTCTCGCGCGCGCGCGCACACAGGGAGCGACCGTTGT 75042
OY 1786 ccaactgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1845
Db 75043 GCACTGGATGACACAGCTGCTTACGACGTCGAGGGGCTCGAGCGCGCGCGCTGGCGCT 75102
OY 1846 catccaag 1854
Db 75103 GCTCAAGAG 75111

RESULT 5
AB048713 AB048713 3277 bp mRNA PLN 11-JUL-2001
LOCUS Pisum sativum PSSCR mRNA for SCARECROW, complete cds.
ACCESSION AB048713
VERSION AB048713.1 GI:13365609
```

```
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 3277
/organism="Pisum sativum"
/cultivar="Alaska"
/db_xref="taxon:3888"
/clone="PSSCR1 and PSSCR2"
/clone_1ib="lambda GEX5 pea shoot cDNA"
/dev_stage="5 days old"
/note="PCR-PeasCR41, PCR-PeasCR33"
322..2781
/gene="PSSCR"
322..2781
/codon_start=1
/product="SCARECROW"
/protein_18="BAB3155.1"
/db_xref="GI:13365610"
/translation="MAACALPAGVGGNTPTDETNNSNSTSSNISTEDFHNMPOOP
HHSERKLLRKRMASEMELQHHNNNNNDYHFSRRNTNTSLNCLPATTOKGVTTT
TTTASLNKNNNNNNNNNNYHNNNNNNIINNHNNTVAISROVIAIONPVTVTN
YSTMLPSSCSNLNNSSTSAANYTHOQPIPEONNLPETICGFSGLIPRSONRN
RTNNSSNNRNTVTVDVSSPSMETSTATTWIDGILKDLHTNSYSIRPOLIN
VREITTPCNPIALVLEHRLLEPTNCPERRRNSTESGVAVNVCNVLASVANS
SVKLMNRDADVPTSLHPSDSSTLLNOMNMPPMNGAQIINNPNPVSILVLPQ
PLSTQDOQHOLOQHPEDLAPATTTTSAELALARKKEIKEXOKKDEGLHTLTL
LLOCAEAVSANTLDOANKMLEISOLSPFCTSORAAVSEAI SARLVSSCIGIYA
TLPVSHPHNOKVYASAFOVNGEISPVKPSHTPANAOEAFEREERHIIIDIMO
GLMPGLPHIILASRGGPPYRLUTGLGSMETLEATGRKLSDFANKIGLPEPPVAV
KGVNTDVEKLVNSKEAAYVHMLDHSIDYDGTNTLMLQRLAPRVAVVEEDLSN
AGSELDGRVEAIIHYVSALFDLSGSYGESESEERHVBQOLISREIRNVLAVGFSRS
EIKFHNMRKLEQCGFRGVSLAGNAATQASILLQMPSEGYTLVEDNGIKLQKIDC
LITFASMPHPHTNTIIPHN"
```

```
BASE COUNT 1038 a 665 c 562 g 1012 t
ORIGIN

Query Match 8.2%; Score 287.8; DB 8; Length 3277;
Best Local Similarity 63.2%; Pred. No. 1.8e-25;
Matches 461; Conservative 0; Mismatches 262; Indels 6; Gaps 1;

OY 1133 aagagcgaagaagagcagcgagcagcagcagcagcagcagcagcagcagcagcagc 1192
Db 1606 AAGAGAAAGAGAGATTAAGACAGAAAGAAAGATGAAGAGTTTGACACCTTCTA 1665
OY 1193 acgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1252
Db 1666 ACACCTTCTTCTCAATGCGCAAGACCTGTATCAGCAGAGATCTTAGACACAGCAACAG 1725
OY 1253 acgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1312
Db 1726 ATGTTACTGAGATTTCCGAGCTCTGCACTCATTCGGGACTTGGCGCAACAAAGATGACA 1785
```



```

/db_xref="GI:6983861"
/translation="MPRYKPEYLFGEVGEVQELRTMSFVIGTSEADYVAOIPHDREE
KCRVKTINSDSDIPSYMEAGGRNTHICQCVATIALGELDRASDOLADREYVH
PROPOGSDRGSYLETETENDATTKHLVEMLMAMDSTRATVYLAADREBRNGKICK
LEDAVDLEKELEALAGAPQKARILTRARKALVPRQLKRVKVEKEVAPAOA
DPPVILISDEEESKRNSKIEWAVATQDEDEPSIDLRCLELE"
complement(40024..43033)
/note="5' UTR"
join(45765..45778,48275..48378,48699..48745,49124..49342,
51652..51729)
/note="ESTs A0077636(E30684), A0064227(E20182),
A0082144(E20182) correspond to a region of the predicted
gene.; hypothetical protein"
/codon_start=1
/protein_id="BAA90797.1"
/db_xref="GI:6983862"
/translation="MGISSTRRYKSNPSPSPARRPLPSPACDALTPDLRSRSGK
PNOKFRTPOQVGLRCSSTYLSWSDSPSPSPISHRPLGGRSRKRPATMTYRMS
LSTSVYMGVATAGLIGIFLFGKEKEFQNYLCRGEERLQODRAAMGN"
complement(join(54820..57727,58608..58723))
/note="Similar to gene for Plb (AB013448)"
/codon_start=1
/protein_id="BAA90798.1"
/db_xref="GI:6983863"
/translation="MAETALSMARSMLGISAASAAGEMSLNGVKEIMFNDE
LKTMOAFLIAETMKKDLKYAEQVRSLSYDIECLEEFVAVHGNOSLQOLNL
KDRRIKAVKIRNLKLEYSRNTRYNSIKMANNTFDEISEMEDVNSRINDEA
KLNGEDPRKEELDKIMDANDDDHCVMGVGGIGKTTVARKIPESEKDIINFP
HRMAYVSQFSMIEMKDISOLIGESIKREGPRIRAHDTGTLRDLGKLTRECV
VFDMATDITWENTREFALPSKNKRSRVLTTRLGVAACATGPRVRLKLTRECV
AIDLLEKMESEKEDKNDNLKSIYQILVKKCGCLYAVTIGAMFANPSSKWEEM
CROLPELENNPSPCYEALIRVYTLISGLHPSHLKCYLYLSIFPEDIEIKRHILNR
WAEGLVARAVGMTISDVESEYFDELISRGNSIOPSRNMEGHSVCRVADIMDIYS
ISKRENVYSTGDNVSTVIEKFRHLSCHGNSPIYMGDSRVSRLVGEPEOPRML
VGSICSAOPFMLRYDLNFAVSQKXKJNKJLRLHYLTHMRRSSTYALPSS
IGKLONLQVDIRSEISTPTDISKLMRIIRCSKGPYFTFFDPDEPIKCLHT
LRMPLMLPLVGSANRYTIAELHRAVSHMSTPGVPTGISKLQVLEVDLKL
LTKSKAIEELGEMLLQKLRVSTKGADKKRKLCAIEKLSLQSLCYEAGYFTTG
TLEWLGPCNFPPLRLKLELHGRIRVMPDSPFNILKOLKIFRLSELDRALIEIIGT
LPNMLNLNGRAVGNELAFKKHEPNLKEKELISMSLSEGRFRFERKDLPHMEREI
ECCOLRSGIVIKHLEQNLNEISGCDVAGDLEOFLKPAKPRPALRLDQDRKIDLG
LPVYLTDESNNEEAKESVHDADGESSVISSP"
join(59365..59367,59518..59647,61034..61132,61435..61862,
61902..62036)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA90799.1"
/db_xref="GI:6983864"
/translation="MKSMTFVCKRVLIDGKVSGLKNFTKLGONELDODINDLNICA
DPLPIKGISETDLITARSTVSTNTYDLKKGPMASPANSCPAGEILLPHGELIH
ARRRAPPCLRAPDAHAGSTWTSARPTGRSIIAANGSGRATVALTRGGRGAA
EVALVLLATVALGAMSLAVRVIMHMGVRLGADAPGDEPSECYJHHLRRLSDVDPG
RORLISATAPPOPSSTELLALARNDGRTGYSRNENRPRGM"
join(63578..63833,66169..66197)
/note="EST D15612(C0952) corresponds to a region of the
Query Match 6.9%; Score 241.6; DB 8; Length 176979;
Best Local Similarity 57.6%; Pred. No. 8,9e-21;
Matches 515; Conservative 0; Mismatches 364; Indels 15; Gaps 4;

```

```

Db 174751 GAACGGCGTCCGCCGCCGAGGAGGAGCGGCTGAGACTGTCCGCCGCT 174692
Qy 1201 gctgcagtcgcccgaagccgctgaagcggacaactcgacgcgcgcacgaagcgtct 1260
Db 174691 CACCGGCTCGCGCGACTCCCTCCCGCGGACCAACGAGGCGCCACTACTCTGGC 174632
Qy 1261 ggaagtcgcggaactagcagccgtctcg---caactgcagcagcgctgagcccta 1317
Db 174631 CCGGCTCGCGGAGATGGCTCCGCGGCGGCCACACCGCGATGACCGTGTGGCGCTA 174572
Qy 1318 ctgcggaagccatgctgcgcgcgcgtcgcaactctgcctctgagcgtt-----acgc 1371
Db 174571 CTTCACGAGAGCGCTCGCGCTCCGCTCGTCGATGTGGCGACATGTTGACATCGG 174512
Qy 1372 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1431
Db 174511 CCGCGCGGAGACTCACGACGACGCTTCGCGGCGGCGACGACGACGACGACGACGCT 174452
Qy 1432 ccagtgctcaagcagcagccctctcgaagtcctcgaactcgaactcgaactcgaactcga 1491
Db 174451 GCGGATACCTAACGCGCATCAGCGCCATCCCGAGGTTCTCGACTTCACGCTCACAGAGG 174392
Qy 1492 catcagaagcgcttcgagcggaggaagcgcgctgcaatcagatcgacatcgatcgca 1551
Db 174391 CCTCTCCCGCGAGTTCCAGGCGGACGAGCGCGCTCCACGTCACGACTTCGACATCAAGA 174332
Qy 1552 gggcgcgagtcgcccgcgcgcctctcgaactctgctcgcgcgcgcgcgcgcgcgcgcgc 1611
Db 174331 GGGGCTCCATAGCGCGGCTGCTCCAGAGCTTGCGCGCGCGCGCGCGCTCCCGCGCA 174272
Qy 1612 ggtgaagcgtacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1671
Db 174271 CGTCCGATCACCGAGTGGCGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCT 174212
Qy 1672 ctgcgaattcgccgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1731
Db 174211 GGGCGCGTCCGCGCGCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 174152
Qy 1732 cggcaatgttgacccggaagaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 1791
Db 174151 CGAGGACGTCGCGCTGTGATGCTCCACGTCACAGCGCGCGAGTGTGCGTGAACG 174092
Qy 1792 gctgcacacatcgctcgaagcgtcgaagcgtcgaagcgtcgaagcgtcgaagcgtcga 1845
Db 174091 CGTCTCGCATGACCGCGCTGCTCGCGAGCAGCGCGCGCTGACGACTTCT 174038

RESULT 8
AP003259 174142 bp DNA HTG 21-FEB-2001
LOCUS Oryza sativa chromosome 1 clone P0466H10, *** SEQUENCING IN
DEFINITION
ACCESSION AP003259
VERSION AP003259.1 GI:13027289
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0466H10.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaralioideae; Oryzaceae; Oryza.
REFERENCE 1 (sites)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GAS) genomic DNA, chromosome 1, PAC
clone:P0466H10
JOURNAL Published Only in Database (2001) In press
2 (bases 1 to 174142)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agricultural Resources, Rice Genome Research Program; Kannonaki
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/,

```


[illegible][illegible]

Db	917	TCCTCGACGGCCGCTTCGCCGACCTCCTCAGCGGACATTACAGAGTCCT-----	967
QY	1452	gcccctctgcgaagttctctgcacattaccgcgcacaccagcgcatccaggaagcgcttcgaagc	1511
Db	968	GCCCTTAACCTTAAGTTGGCGCACTTACCGCCACACGAGCCATCTCTGAAGGGGTTTCGGCG	1027
QY	1512	ggagagagcgcgctgcacatcatcagacctcgacatcatgcaagagggcctgcagttgacccgagc	1571
Db	1028	GCTCCGCCCGCGTGCACAGTGTGCAGCTTCGCGCATGAAGAGGGGATGACAGTGGCCCCGAC	1087
QY	1572	tcttcacatctcttgtctcccgcccgaggcgccgcgccaggtatgagcttaaccgagcctcg	1631
Db	1088	TTCTCCAGGCCCTTCGGCTTCGCCGGCGGCCGCTTCCTCTGTTCCGCTCACCGGCGTCGG	1147
QY	1632	g-----ggcgtccatctgagagcgctcagagcccaagggagagacgctctccagatt	1679
Db	1148	GCCCCCGCAGCGGAGACGAGACGAGCGCCCTGAGCAGAGTGGGGTGGGAAGTCGCCAGT	1207
QY	1680	tctgcagacagctcgcgcctgccttcgaattctctgacgcgtgcgcg	1724
Db	1208	TCGGGACACACATCCGGGTGAGATTCCAGATGACGGGCGCTTCGTGC	1252

RESULT 14
CEY39B6_3
WPCOMMENT
Sequence split into 4 fragments LOCUS Cey39B6 Accession 295399

Continuation (4 of 4) of CEY39B6 from base 300001 (Z95399 *Caenorhabditis elegans* chromosome

Query Match	4.98;	Score 170.8;	DB 2;	Length 56870;
Best Local Similarity	48.18;	Pred. No. 3.2e-12;		
Matches 484;	Conservative	0;	Mismatches 522;	Indels 0;
			Gaps	0

[illegible][illegible]

RESULT 15

LOCUS	AB017067	83689 bp	DNA	PLN	27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MCC20.				
ACCESSION	AB017067 BA000015				
VERSION	AB017067.1 GI:3510343				
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_1lb:Mitsui PLN				

ORGANISM

REFERENCE
AUTHORS
TITLE

Elkayova: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eustosids II: Brassicales: Brassicaceae: Arabidopsis.
1 (sitee)
kaneho, T., Kato, T., Sato, S., Nakamura, Y., Asamizu, E., Kotani, H.,
Miyajima, N. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 5. IX.

JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL	DNA Res. 6 (3), 183-195 (1999) 99397451 2 (bases 1 to 83689) Nakamura, Y. Direct Submission Submitted (26-AUG-1998)	Kazusa DNA Research
--	--	---------------------

COMMENT

For the latest information on annotation of this cluster, please see http://www.kazusa.or.jp/xaos/cgi-bin/abd_graph.cgi?c=McJ20. Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),

FEATURES

Source

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremlini1.zool.iastate.edu/cgi-bin/sp.cgi>) Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K16122 and the 3' clone is K5014.

Location/Qualifiers

```

1..83689
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MJC20"
/clone_id="Mitsui P1"
complement(join(393..464,554..643,717..822,901..981,
1054..1265))
/note="gene_id:MJC20.1
similar to unknown protein
sp145966"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08424.1"
/db_xref="GI:9757936"
/translation="MDOTONTAGIGINGSTIKETINATVYDSEKNIQIINSNOK
ILELHOLKTVSFTTPASOLHLOLRINSYVMELDNNAKISDCNTQVPEIYNLDD
GKNDEFTKVLNKNCTAKNQVTKGSKDARKGLKLELEAPDEVDYRDIRAS
YAAEAKRLAQTSVLPNGDAKRVSEL"
2215..3432
/note="gene_id:MJC20.2"
/codon_start=1
/evidence=not_experimental
/product="SCARECROW gene regulator-like protein"
/protein_id="BAB08425.1"
/db_xref="GI:9757937"
/translation="MTYKRIDRLDPSDDPSAKRRIEPPEETLENDGAAILKLSLL
LOCAEYATDHLRASTLSEISEICFSGSPERVAVYPAQALOTRVISYSGACS
PLSEKPLTVVOSOKIFSALOTYNSVSLPIKFSHTNQAIFQALDGDPSVHIIIDLVY
OGLQMPALPHILASRPKRLSIRITGCGSSDLASGRAPASINIPPEPHIE
GITGNLIDPSQLATROGEAVVHMOMRLDVGNNLETLLEIRLKPMLITVBEEL
SYDGGSGFLGRFEVALHYSALEFDALGDGEGSGERTVEQTLCTENIVAHGG
RRKRMKKEELSRVGRFVRPLRGNPATQALLGLMPMNGYTLVEENGTLRLEMKDLS
LTPASAKSQPPD"
4272..4568
/note="gene_id:MJC20.3
unknown protein"
/codon_start=1
/evidence=not_experimental
/product="GTPase activator protein of Rab-like small
GTPases-like protein"
/protein_id="BAB08427.1"
/db_xref="GI:9757938"
/translation="MMFRLRLTKVFRSPILAKWIVFADGSSGGVGRNGYIGTMS
SAVTPSVGLAVTAVMAVAFTVYSRRTYHNDLIVLIDFKAKKQCLEFT"
join(5257..5345,5552..5687,5778..5827,6222..6275,
6493..7684)
/note="gene_id:MJC20.4"
/codon_start=1
/evidence=not_experimental
/product="GTPase activator protein of Rab-like small
GTPases-like protein"
/protein_id="BAB08427.1"
/db_xref="GI:9757939"
/translation="MIFCLLVILIGISWTVCSDENVYKLSRGISGSPMSRRKR
VLQPKQNAFTBEGRLSDGVKFLKKVRSRGVHPSIRPEVWFLGVLDLQKSKER
DSIRQLTLEYENLRQCRERHERYENVCDSKTQASNTEDSQVLESHDIIEVESST
RTIVDESEKLSNESIMQDENCEKSDVETDAAGNDSSTNPETETSPLEAVEES
HNTVQKEISSPSKRQADEDEMTWQRIIRLDAYRANDEWYSPLOAVSET
KARGLAIOVGINDYDLHLEPCRFHARLVGLIAYAVYDPEIGYCOGMSPLSLAY
MEDVLAEMCFGVGSKARHNPRLDVGIRGROISMKIKKEDILHYLHLEMLBED
CFYVYRVVYVLFRELTFEOTLCLMEVMAQDAIRTGAKATWGRIRLRAPTEDL
LYAIAASVIOARRKTIIEKYSMDIMKECMSMAGHLDVWKLLDADHLYVNVHDKI"

```

CDS

```

join(8058..8569,8692..8811,9050..9119,9477..9590,
9713..9796,9924..9974,10134..10163,10275..10337,
10430..10495,10589..10717,10851..10886,10966..11085,
11163..11354,11483..11593)
/note="gb|AAFG32311.1
gene_id:MJC20.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08428.1"
/db_xref="GI:9757940"
/translation="MADVEEPOLONGAPAESETEONPIPEPOLQTEPKLTGEIPI
EADLIPEVQOSEVTDKPEVEQSEKPEVTVYVDAPPEAQSEVKEEQSVVTPR
KPDLDVDSLSPGSEIEPIRSTVEQESTVYLVKDDGKFTFMRLELSKEEGD
CPHSDASPESESAPENNPAMDILNRIQVNDDEGRSQRVLAFAKRYASAIER
NPDHDLALYNALIIQESADNVPSPVSPKDDLLIEACKRYDEATRLCPLDYAYN
WALAISDRKIRGTREKAEDELMEQADYERAKVOLNMNSQALNMKGVLQDELQIYP
AREKERVTAITSKFAIRLQDFDHRATYNGVLYLADTLTLTGSGGKDMPPG
ELYSQSAIYIAAHSILKPSYVSASSALRYSMLPLPLKGVLYAPVGNLSLPHSD
WKRTFEELNHERLQVLKPEPEMRNLISGRAEMTSTVERKTVKVNITEISVTPCA
DLTPPGAGLCIDTIGHGVFLVADSMESLDGMLAIRLVYTIYARGSVDVLAITG"
join(11993..12130,12206..12427,12522..12601,12703..12778,
12871..12961,13075..13101)
/note="gene_id:MJC20.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08429.1"
/db_xref="GI:9757941"
/translation="WASLSCLPICGLSSASLYTKARLLSSSSSPSPFLAYPLQNP
MKLCKIERVICRAEFSDAPLVTAIGACIISSFFPAKRVNDEEENSALVSTD
NMLAMGIISFIEPIFNMLSWFAMLDTRKRSYAVVALYVLYLSNISIPESMLP
ITSYVIGIIVHVLDEASIANGVETLAFPRNSDDPSSKRIHKKHPRKGNSDEN"
join(13512..13697,13866..14012,14098..14196,14285..14476,
14612..14731,14818..14869,14970..15194)
/note="gene_id:MJC20.7"
/codon_start=1
/evidence=not_experimental
/product="GAMM1 protein-like"
/protein_id="BAB08430.1"
/db_xref="GI:9757942"
/translation="MAPPSSVVRVYSTATSPSPSTSYKVGCTHNGSFHDEALCFMI
RLVDFSGADIVYRSKDPKTLAEELDVLVGVGYDPEHRYRHQKFEVYGHGPNRK
LSSAGLVYKHPEKELIAKLEVQDHPVLRFLAVYKSFMEADVANGINRYDQDQ
PFRVYNNTHLSRVRGLNDWIDPQOSCKENEFQRMALAGFELESVOFHARSM
PARSIYMOCLERFKTDPGSEIMILDRCEPMKRLFELEQMKIEPLIKYIYODERA
KOMRYOAVAVAPDRFENRKLPEKMRGLRDELSAAATIPGCVFHMGGFICGNQSD
GALSMAGTALRT"
15998..17122
/note="emb|CAB72466.1
gene_id:MJC20.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08431.1"
/db_xref="GI:9757943"
/translation="MEISGEEDKEEAVTLPREVSKISISDGNKEFYQIILNPNQCFE
NFRMDKPEYKICDLQTRGLRHTNRKIAEOLAIFLEIIGHNLRTAAVOLECYSG
ETISRHPNNVNAIVATISKDFEOPNSNDTELENDPDKGVGVDSDHPIPVMGVDE
OCPFRNGCLITONVLAASSFDLRPNRYLACBEGSASQOVLNMLHTRRKLQYPOCK
YIIVDNKYNLPGFLAPYHGVSTNSREKAKEMFNRHLLHRAHLRTFGALKEKFPIL
LSAPPYPIQOVKLVIAACALHNVLRLEKPPDLVFRMEDEETLAEAGDEKVALLEEQ
VIVQOEHGFRPEEVEDSLRLDEIASLMMHYVQNMST"
complement(join(17318..17892,17980..18361,18437..18594,
18670..18890,18991..19218,19303..19340,19706..19770,
19981..20140))
/note="gene_id:MJC20.9"
/codon_start=1
/evidence=not_experimental
/product="MAP kinase"
/protein_id="BAB08432.1"
/db_xref="GI:9757944"
/translation="MEGFNRIRKRRHNDGRDRREICKSEMGIISFAKLRFRSFRPT

```

CDS

CDS

CDS

CDS

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2002, 15:36:30 ; Search time 298.88 Seconds

(without alignments)
10068.294 Million cell updates/sec

Title: US-09-265-585c-95

Perfect score: 3510

Sequence: 1 cgcgtacgtcagcctactca.....tagccatcaagcaagtattta 3510

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3510	100.0	3510	21	Maize Scarecrow nu
2	698.4	19.9	2151	18	Maize ZCARECROW ZC
3	698.4	19.9	2151	21	Maize ZCR gene par
4	358.4	10.2	2163	18	Arabidopsis SCAREC
5	262.4	7.5	2161	21	Arabidopsis SCAREC
6	218.4	6.2	969	21	Maize CBPBT14 par
7	172.4	4.9	2709	20	Composite DNA sequ
8	171.8	4.9	1768	20	Wheat Rht clone C1
9	171.8	4.9	2125	20	Wheat Rht clone 5a
10	169.4	4.8	1369	21	Arabidopsis thalia
11	169	4.8	1746	20	Consensus cDNA seq

12	157.4	4.5	1328	21	AAC39840
13	155.8	4.4	2255	20	AA36280
14	140.2	4.0	1337	20	AA217263
15	131.2	3.7	114955	20	AA53491
16	130	3.7	321	18	AAT95765
17	130	3.7	321	21	AAC65296
18	127.4	3.6	377	20	AA36268
19	122.4	3.5	332	20	AA36255
20	118	3.4	1000	21	AAA02484
21	116.2	3.3	114955	20	AA53491
22	108.6	3.1	1218	21	AAA02488
23	107.8	3.1	3957	22	AA09686
24	106.8	3.0	1593	21	AAA02504
25	106	3.0	8438	15	AA073500
26	101.8	2.9	1459	21	AA02528
27	99.8	2.8	1359	20	AA217254
28	98	2.8	1127	21	AA02477
29	96.8	2.8	100	21	AAC65284
30	95.2	2.7	100	18	AAT95769
31	94.4	2.7	1779	22	AAF25480
32	94.4	2.7	1779	22	AAF25481
33	93.2	2.7	1523	20	AA217507
34	93.2	2.7	1523	20	AA217479
35	90	2.6	1925	20	AA90924
36	90	2.6	5120	22	AAC84677
37	89.4	2.5	320	21	AAA38183
38	88.8	2.5	1126	21	AAA38185
39	88.4	2.5	1126	21	AAA02538
40	88.4	2.5	6530	14	AA051557
41	88.4	2.5	6530	20	AA222072
42	88.4	2.5	6530	22	AAC68948
43	87.2	2.5	1926	21	AAA50254
44	87.2	2.5	1926	22	AAF82902
45	87.2	2.5	2580	21	AA75454

ALIGNMENTS

RESULT 1
ID AAC65299 standard: cDNA, 3510 BP.
AC AAC65299;
DT 09-FEB-2001 (first entry)
XX
DE Maize Scarecrow nucleotide sequence.
XX
KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
KW transgenic plant; cell division; molecular marker; herbicide resistance;
KW salt resistance; pathogen resistance; insect resistance; ss.
XX
OS Zea mays.
XX
PN WO200053723-A2.
XX
PD 14-SEP-2000.
XX
PF 07-MAR-2000; 2000WO-US05875.
XX
PR 10-MAR-1999; 99US-0265585.
XX
PY (UYNV) UNIV NEW YORK STATE.
XX
PA Bentley PN, Di Laurenzio L, Mysiocka-Diller J, Malamy JE, Pysch L,
PI Helariutta Y, Bruce W, Lim J;
PI WPI; 2000-594315/56.
DR P-PSDB; AAB28595.
XX
PT Scarecrow gene useful for producing transgenic plants expressing genes
whose product increases starch, lignin or cellulose biosynthesis and

PI Pysh L, Wysocka-Diller J;
 XX WPI: 1997-549683/50.
 DR P-PSDB: AAM38192.
 XX
 PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
 PT cell division and therefore alter root development, or alter plant
 PT stem or hypocotyl gravitropism
 XX
 PS Claim 6; Page 137-138; 221pp; English.
 XX
 CC This sequence comprises a cDNA clone of the maize ZCR gene, a
 CC homologue of the Arabidopsis SCARECROW (SCR) gene (see AAT95753).
 CC It was isolated using a PCR-cloning strategy (see also AAT95773-78).
 CC It encodes ZCARECROW (see AAM38192), which shows homology to
 CC Arabidopsis SCR (AAM38178). SCR is expressed specifically in embryo
 CC root progenitor tissue and in certain root and stem tissues. It
 CC regulates a specific asymmetric division, and controls gravitropic
 CC response in aerial structures and root formation. Nucleic acid
 CC molecules (see AAT95753-66) encoding SCR protein sequences (see
 CC AAM38178-201), SCR proteins lacking 1 to 4 of MOTIFs I to VI, or
 CC SCR MOTIF 1, II, III, IV, V or VI are claimed. Transgenic plants
 CC can be engineered to overexpress SCR, so that cell division is
 CC increased in roots, resulting in thicker root development, while a
 CC plant with an altered stem or hypocotyl gravitropism is less
 CC susceptible to lodging. Plants that contain an antisense
 CC molecule that suppresses the expression of endogenous SCR gene
 CC product show thinner root development. A gene of interest can be
 CC placed under control of a SCR promoter and expressed in a plant to
 CC confer herbicide, salt, pathogen or insect resistance, or when
 CC expressed in stems to increase starch, lignin or cellulose
 CC biosynthesis (all claimed).
 CC
 XX Sequence 2151 BP; 492 A; 558 C; 523 G; 578 T; 0 other;
 XX
 Query Match 19.9%; Score 698.4; DB 18; Length 2151;
 Best Local Similarity 91.4%; Pred. No. 1.9e-104;
 Matches 770; Conservative 0; Mismatches 51; Indels 21; Gaps 2;
 QY 1112 gcagagatgtctgcgcgcgaagagagagagagagcgcgcggaacagcgagac 1171
 DB 869 gcgcctcgcgcgcgcgcgaagagagagagcgcgcggaacagcgagac 928
 QY 1172 gaggagggcctccaccc-----tgctgacgtctgtcgtcagctgcgcgcgcgcgaa 1224
 DB 929 gaggagggcctccacccgtctgagtgctgacgtctgtcgtcagctgcgcgcgcgaa 988
 QY 1225 cgcgcgaacacttgacgcgcgcacacagacgtctgtgagatcgcgcgcgcgcgcgc 1284
 DB 989 cgcgcgaacacttgacgcgcgcacacagacgtctgtgagatcgcgcgcgcgcgcgc 1048
 QY 1285 gttcggacactcgc 1344
 DB 1049 gttcggacactcgc 1108
 QY 1345 cgtcagctcgtcgtgc 1404
 DB 1109 cgtcagctcgtcgtgc 1168
 QY 1405 cctcagcgccgc 1464
 DB 1169 cctcagcgccgc 1228
 QY 1465 gttcgcgactcgc 1524
 DB 1229 gttcgcgactcgc 1288
 QY 1525 gcaacatcgc 1584
 DB 1289 gcaacatcgc 1348
 QY 1585 tgctcccgcccgcgcgccgc 1644

DB 1349 tgctcccgcccgcgcgccgc 1408
 QY 1645 ggcgcctcgc 1704
 DB 1409 gcgcctcgc 1468
 QY 1705 cgcgtctcgc 1764
 DB 1469 cgcgtctcgc 1528
 QY 1765 gaggcgagagcgccgtcgc 1824
 DB 1529 gc 1588
 QY 1825 cgcctcgc 1884
 DB 1589 cgcctcgc 1634
 QY 1885 gactctcgtcgc 1944
 DB 1635 cctcgccttcttcacatgccaatcttgatgcaatcattgacacatttcagctgctgaca 1694
 QY 1945 ct 1946
 DB 1695 tt 1696
 RESULT 3
 AAC65298
 ID AAC65298 standard; cDNA; 2151 BP.
 XX
 AC AAC65298;
 XX
 DT 09-FEB-2001 (first entry)
 XX
 DE Maize ZCR gene partial nucleotide sequence.
 XX
 KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
 KW transgenic plant; cell division; molecular marker; herbicide resistance;
 KW salt resistance; pathogen resistance; insect resistance; ss.
 XX
 OS Zea mays.
 XX
 PN WO200053723-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 07-MAR-2000; 2000MO-US05875.
 XX
 PR 10-MAR-1999; 99US-0265585.
 XX
 PA (UYNX) UNIV NEW YORK STATE.
 XX
 PI Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L,
 PI Helariutta Y, Bruce W, Lim J;
 XX
 DR WPI: 2000-594315/56.
 DR P-PSDB: AAB28591.
 XX
 PT Scarecrow gene useful for producing transgenic plants expressing genes
 PT whose product increases starch, lignin or cellulose biosynthesis and
 PT confers herbicide, pathogen or insect resistance
 XX
 PS Claim 6; Fig 17A; 200pp; English.
 XX
 CC The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MOTIF III (VHID) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell

CC confers less susceptibility to lodging in the transgenic plants than a CC wild-type plant. SCR gene sequences are also useful as molecular markers for a quantitative trait e.g. root or gravitropism trait in molecular breeding of crop plants.

SQ Sequence 2161 BP; 569 A; 668 C; 312 G; 606 T; 6 other;

Query Match	7.5%;	Score 262.4;	DB 21;	Length 2161;
Best Local Similarity	55.0%;	Pred. No. 4.9e-34;		
Matches 610; Conservative	0;	Mismatches 477;	Indels 22;	Gaps 4

[illegible]

Db	1394	gctctcccaattacacagataaactccctccctcttcaactctgccttaccctgag	1453
QY	1727	aagcgcgcgaatgcttgaccgcggagaagactagagggtacagagcgggagccgtgcgcgc	1786
Db	1454	aaagttcgaacttcgcactgtagagactaaatcctagcaaaacccaactctccctct	1513
QY	1787	cactgcctgcacccactgcgtcctacgaagtaactgctgcgaactcaaaagctctgctc	1846
Db	1514	cactgccttcaacatctctcttaatgtaagtaactccctctgacgacaaactctctctga	1573
QY	1847	atccaaagtgtagagagagtaaccatct	1875
Db	1574	ctccaacgctaaataaacattacccttt	1602

RESULT 6
 ID AAC65327 standard; cDNA; 969 BP.
 AC AAC65327;
 XX
 XX AAC65327;
 XX
 DT 09-FEB-2001 (first entry)
 XX
 DE Maize CBPPT44 partial cDNA sequence.
 XX
 XX SCARECROW gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
 KW transgenic plant; cell division; molecular marker; herbicide resistance;
 KW salt resistance; pathogen resistance; insect resistance; ss.
 XX
 XX Zea mays.
 OS
 XX WO200053723-A2.
 PN
 XX
 PD 14-SEP-2000.
 XX
 PF 07-MAR-2000; 2000WO-US05875.
 XX
 PR 10-MAR-1999; 99US-0265585.
 XX
 PA (UYNY) UNIV NEW YORK STATE.
 XX
 XX Benfey PN, Di Laurenzio L, Wyszoka-Diller J, Malamy JE, Pysh L;
 PI Helariutta Y, Bruce W, Läm J;
 XX
 DR WPI; 2000-594315/56.
 XX P-PSDB; AAB28603.
 PT
 PT SCARECROW gene useful for producing transgenic plants expressing genes
 PT whose product increases starch, lignin or cellulose biosynthesis and
 PT confers herbicide, pathogen or insect resistance -
 XX
 PS
 PS Example 8; Fig 33; 200pp; English.
 XX
 XX The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MYR1 III (MYR1D) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryos and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.
 XX
 XX Sequence 969 BP; 133 A; 333 C; 360 G; 143 T; 0 other;

PT New Tricum aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX
XX Disclosure: Fig 3a; 88pp; English.

	RESULT	8
AA36278		
ID	AA36278	standard; cDNA; 1768 BP.
XX		
XX	AA36278:	

CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence represents
CC the wheat Rht clone 5a1 genomic sequence.
XX

SO Sequence 2125 BP; 325 A; 768 C; 723 G; 309 T; 0 other;

Query Match 4.9%; Score 171.8; DB 20; Length 2125;
Best Local Similarity 58.4%; Pred. No. 2,1e-19;
Matches 377; Conservative 0; Mismatches 232; Indels 36; Gaps 3;

```
OY 1092 cggcgccgacgcgcgcacaaagcagcagctgtctgcccgcgaagagcgagagagagc 1151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 632 ctgtgtgtgaggtctgcccgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 691
OY 1152 agcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 692 tctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 751
OY 1212 ccgagcgccgtgaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 752 cggagcgccgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 811
OY 1272 agctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 812 tgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 871
OY 1332 tgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 872 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 916
OY 1392 ccgagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 917 tctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 967
OY 1452 gccctctgtcaagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 968 gccctctcaagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1027
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1512 gggagagcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1028 gctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1572 tcttcacacatctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1088 tcttcacagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1632 g-----ggcgttcacatgagagcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1148 gcccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1680 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1208 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 10
AAC49468
ID AAC49468 standard; DNA: 1369 BP.
XX
AC
XX AAC49468;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61273.
XX
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX

```
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
```


[illegible]

RESULT 11

ID	AAK36285	standard; cDNA; 1746 BP.
XX		
AC	AAK36285;	
XX		
DT	16-JUL-1999	(first entry)
XX		
DE	Consensus cDNA sequence of wheat Rht clone C15-1	

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
 KW paclobutrazol; ss.

05 *Triticum aestivum.*

PN W09909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998;

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum Aestivum polynucleotides - encode a polypeptide which

PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 2a; 88pp; English.

The specification describes polypeptides encoded by the Rht gene (and its homologues) that, when expressed in *Triticum aestivum*, inhibit growth of the plant. This growth inhibition is antagonised by gibberellin. The products can be used to provide Rht expression in plants, conferring a dwarf phenotype on a plant which is correctable by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made more resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol e.g., to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence represents the consensus cDNA sequence of wheat Rht clone C15-1.

50 Sequence 1746 BP; 375 A; 590 C; 495 G; 278 T; 8 other;

Query Match	4.8%;	Score 169;	DB 20;	Length 1746;
Best Local Similarity	61.0%;	Pred. No. 5.8e-19;		
Matches 346;	Conservative	0;	Mismatches 185;	Indels 36;
			Gaps	3;

[illegible]

RESULT 12

ID AAC39840 standard; DNA; 1328 BP.

AAC39840;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26094.

KW Hybridisation assay; genetic mapping; gene expression control;

metabolic pathway; promoter; termination sequence; ss.

OS *Arabidopsis thaliana*.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140941.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

[illegible]

Search completed: April 22, 2002, 18:00:24
Job time: 8634 sec


```

Db 39283329 GTCAACCGTGGCGGCGGTGGCTTACCOCGGTGTCATTTCAGGCGCGCTTTGGCCGCTTAGGC 39282710
OY 1403 cgcctccacagcgcgcgttcgcgcgcgcttccaggctgtttcaacagcatcaagcccttgatc 1462
Db 3928269 GCCCCTGGCCGCCGGCACACACTTGTGGCCGCGTCAATGCTGGGGGTCCCGCGTGCGCGCGC 39282110
OY 1463 aaattctcgacttaacgcgcgaacaaagccatctcacagagcgcgttcgtagcgagagagcgc 1522
Db 39282109 GCCCGCCGGCCCCCGCGCCCOCGGGTGCCCATCTCCCGCGTCCGGGGAACAACCGCGCGAATCCGCGC 39281510
OY 1523 gtlgcacatcatcgacctgcagaatcatgaagagggcgtcacgttgcgcgggcgtcttccacatc 1582
Db 3928149 GCCCAATTCGGCGTTCGGCGCGTCCCGCGTCCGCGGAAACACCGCGCGAATCCGCGC 39280910
OY 1583 ctltgcctccgcgcccccggggggccccgcacaaggttgaggtctaacaccgcgccttggggcgttcagt 1642
Db 3928089 GTTCCGCGCGCGTCCCGCGCATGCGCGCTTGTGACCGCGCTTAACCGCGCGTGGGTCCGCGAT 39280310
OY 1643 gagcgctctcgaaggccaaggggaagagcgtctctcgaatttcgcagaacagctcgagcctgc 1702
Db 3928029 CAGGGTGGCCGGGGCGCGCGTACACACCGGGCCCCCGCGGTGCGCGCGCCCCCGCGGCC 39279710
OY 1703 ttcgagttctgcgcgcgttcgcgcgaagagcgcgggcaatlgtgaaccgcgagaagctaagggtc 1762
Db 3927969 GTTGTATCCCGCGCGCGCGCGGTGGTCCGCGCGCCCCCTTGGCGCGCTTGGCGCGCGCTC 39279110
OY 1763 acgaagcgagagagcgcgttcgcgcgtctccactcgtctgcacacactcgtctagaagctatg 1822
Db 3927909 AAAACCGGAGAAGCGCCACGCGAGAACCGCTGCGCTTCCGCGCGCGCGTACACCGCCACACCC 39278510
OY 1823 tcgcactccaacacagcgtctggtcatcc 1850
Db 3927849 GCGCGTTTCGCGCGCGCGCGCTTATCCC 3927822

RESULT 3
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFELINGER, F.
APPLICANT: FAIKNER, P.G.
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232.463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109

```

```

      : TELEX: 899149
      : INFORMATION FOR SEQ ID NO: 14:
      : SEQUENCE CHARACTERISTICS:
      : LENGTH: 7218 base pairs
      : TYPE: nucleic acid
      : STRANDEDNESS: single
      : TOPOLOGY: linear
      : IMMEDIATE SOURCE:
      : CLONE: pTZgpt-Fls
      : US-08-232-463-14

Query Match          3.1%; Score 107.6; DB 1; Length 7218;
Best Local Similarity 3.5%; Pred. No. 1.4e-13;
Matches 14; Conservative 269; Mismatches 113; Indels 0; Gaps 0;

QY 103 gccgcgcgcggcgcccaaacagcattgtcttccttacttaaccctatccccgcg 162
    | - - - - - | - - - - - | - - - - - | - - - - - |
Db 1041 GCCTTCAGGTCCAGGAGCGACTTGCATATTTATTTTTTTTTTTTTTTTTTTT 1100

OY 163 ctgggctgcccatgcccattgscgcgcgcgcttcttcttgccgttttagtgacctc 222
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1101 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1160

OY 223 tectctccctctctcttctctctgctctctctccgcgatccacgcccccactcttt 282
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1161 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1220

OY 283 ccccactctatagcacacgcgcacgcctccgcctctctctctaetcttatgtcgccgtg 342
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1221 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1280

OY 343 ccctccccaacacccctccctccgcctctctctcttcttcccaacacacttctctccacta 402
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1281 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1340

OY 403 cctcatcagctagacacacaagaagccgcgcgcgcgcatagttcgcgaagcgcgccgc 462
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1341 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1400

OY 463 gtcgcacatgacctccgcgcgcgcgcgcgcacagt 498
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1401 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTYT 1436

RESULT 4
US-07-945-283-1/C
: Sequence 1, Application US/07945283
: Patent No. 5352596
: GENERAL INFORMATION:
: APPLICANT: Cheung, Andrew K.
: APPLICANT: Wesley, Ronald D.
: TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
: TITLE OF INVENTION: Involving The EP0 and LfT Genes
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Curtis P. Ribando
: STREET: 1815 No. 5352596th University Street
: CITY: Peoria
: STATE: IL
: COUNTRY: USA
: ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 42A
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P

```



```

Db      4185 CTCCTCCGGCTCTCCGGCTACGCAAGCTTCCCCCGCCGCGGAC 4138

RESULT      5
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claite M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match 2.78; Score 94; DB 4; Length 4411529;
Best Local Similarity 45.18; Pred. No. 1.4e-09;
Matches 738; Conservative 0; Mismatches 873; Indels 24; Gaps 9

[illegible][illegible]

RESULT 6
US-09-103-840A-2

QY 508 cctctcc 514


```

Db 3548 GGAGCCACCGCTAGAACCGCTCCGATGCTGTGCACTGATGTTGGACGATCCGCC 3489
OY 703 cgcgtccacgagcgcg-----tgggtgacggcatcatccggaatcatctggga 752
Db 3488 GCGCGCGCCACCGGAGAACCACTCCGACCTGTGTGCGCGCGCCGCGGAATAGCCGCC 3429
OY 753 gcaagcgagcgagcgagcttccatcaacgagcatcatcaacagctccgagatcatcc 812
Db 3428 GCGCGCGCGCGAGCTGAGTTCACCGCCACCTGAGTAGCGCGCGCCACAGCTAGAGCCGCC 3369
OY 813 acccttgacaccgcgcgtcgctcgctcttgagctccgcctccctccctccgag 872
Db 3368 GCTGCTTCACCGCGCGCGGAATAGCCGCTCCGACGTGAGCCACCGCGCCAC-CAG 3310
OY 873 cggaccgagcccgactcgccgagcgccgagcgagcagcatgctctctctgagcg 932
Db 3309 AGTACTTCCCGCTCCGGAACCGCGCGCCGCGCCACCGGAGACTCCCGCTCCAGAGAGC 3250
OY 933 ctccgagcgagctcccgagcgagctccctccctcccgccacgctcccgagacaagc 992
Db 3249 CTCCGCACTAGAGAGCGCGCGCTCTGTATCCGCTCCGCACTGAGAGCGCGCGCGCG 3190
OY 993 ggcgcgaagagatccacgcgcgtgcagcaagcaagcagaggaacgcgcatcgagc 1052
Db 3189 CCGCGGAGAGAGCGCGCGCGAGCTAGAGCCGCGCGCCACCACTCCGAGTACTTGA 3130
OY 1053 cgcagctgc-----ccaagcccgacgagcgaggaagcagcgagcgagcgagc 1104
Db 3129 CGCCCCCAGAGCGCGCGCGCTCCACAGCTGGAGCCACCGCGGAGTAGTGGCGCG 3070
OY 1105 cgcacaacgagcagctgtctgcgcgcgc-aagagcgaggaagcagcgagcgagc 1163
Db 3069 CGCGCACTGAGAGCGCGCGCGCTCGAGTAGCCACCTCCGCACTGAGAGCGCGCG 3010
OY 1164 agcgaggaagagagcgctcactgtctgagctgctgctgcgcgc 1205
Db 3009 CGCCCCAGAGTAGCCACCGCGCGAGTAGAGCACCACCGCGC 2968

RESULT 10
US-08-458-240-1/c
: Sequence 1, Application US/08458240
: Patent No. 6143727
: GENERAL INFORMATION:
: APPLICANT: Roop, Dennis R.
: APPLICANT: Rothnagel, Joseph A.
: APPLICANT: Greenhalgh, David A.
: TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LYON & LYON
: STREET: 611 West Sixth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,240
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/146,930
: FILING DATE:
: APPLICATION NUMBER: 07/876,286
: FILING DATE: April 30, 1992
: APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
: FILING DATE: October 29, 1993

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 204/152
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ. ID NO. 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6530 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match 2.5%; Score 88.4; DB 3; Length 6530;
Best Local Similarity 46.3%; Pred. No. 1.5e-09;
Matches 519; Conservative 0; Mismatches 581; Indels 22; Gaps 6;

OY 104 cccgcgagcgcccaaacagcaactgtccctccctcaatccatccatcccccgcgc 163
Db 4087 CCTTCCACACTTCACAGAGCGCGCCACGAGAGAGCCGCTCCGACCTGAGCCACCGCT 4028
OY 164 tgggtgcgcgagatcgcatatgcgagcgagcttctctctgtgcgtttcagatgagctcct 223
Db 4027 CGGCTTCGAGTAGTACTCTCTCCGACGACCGGCTGCTCCGAGAGAGCCACTCCG 3968
OY 224 cctctccctctctctctctctctctctctccgcgcgacatccacgcccccatctctc 283
Db 3967 CAGCCACCGCGCGCTGCGCGCGGAGAGCGCGCCACAGACCCAGCGCGCTCCGAGTAG 3908
OY 284 cccactctatgcacacgcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 343
Db 3907 CCTCCACCGAGCGCGCTGCTGCGCCACGACGCTGAGTAGAGCCACCGCGCGCA 3848
OY 344 cctccccaacacctccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 403
Db 3847 CCGCGCCAGAGAGAGGACCACTCCACAGCTACAGCTTCGCGGAGAGAGGCGCTCCGTAAG 3788
OY 404 ctccatcagctagaccacaagaagcgcgcgcgcgcccaatgctccgcaagcgccccgcgcgc 463
Db 3787 CTCTGCTGGGGGCGCAGAGAGTGTACTGCTGTGAGAGAGTAATAGCCCCCTCCG 3728
OY 464 tcgaaatgacctcccgcgcgcgcgcgcgcgcaagtcagcgagacctccgagctcagctcgc 523
Db 3727 CGCCAGAGAGAGCGCGCGCGCTGCTGCGCCACGAGCTGAGTAGAGCTGAGAGCTGAG 3669
OY 524 ggcgcgcgtgcgcgcgcgtgttggtgtagtgagcgcgcgctctccgcagcg-cgcagct 582
Db 3668 GCTGCGCGCGCAACTGGAGCCACACACAGAGAGAGCGCGCTCCGAGCTAGAGCTTCGCGC 3609
OY 583 gcccgcgtgcgcacccagctccacaagctgcgcccgcgctccagcaacgagcgcgga 642
Db 3608 CCGGAGTAGCGCGCCACCGCACTGAGAGCTCGCGCGCGGAGTAGCGCGCCCGCACACT 3549
OY 643 ggtgagcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 702
Db 3548 GAGGAGCAGCGCTAGAACCGCGCTCCGTAAGCTTGAGCACTGATATCTGTTGGAGGATCCGCC 3489
OY 703 cgcgtccacgagcgcg-----tgggtgagcgatcatcccgagatcatctggga 752
Db 3488 GCGCGCGCCACCGGAGAACCACTCCGACCTGTGTGCGCGCGCCGCGGAATAGCCGCC 3429
OY 753 gcaagcgagcgagcgagcttccatcaacgagcatcatcaacagctccgagatcatcc 812
Db 3428 GCGCGCGCGCGAGCTGAGTTCACCGCCACCTGAGTAGCGCGCGCCACAGCTAGAGCCGCC 3369
OY 813 acccttgacaccgcgcgtcgctcgctcttgagctccgcctccctccctccgag 872
Db 3368 GCTGCTTCACCGCGCGCGGAATAGCCGCTCCGACGTGAGCCACCGCGCCAC-CAG 3310

```


[illegible]

RESULT 14
US-09-130-114-1
; Sequence 1, Application US/09130114
; Patent No. 5075007

```

: APPLICANT: Horlick, Robert A.
: APPLICANT: Dornaj, Bassam B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: TITLE OF INVENTION: From Multiple Transfected Episomes
: FILE REFERENCE: 0867/1D903J51
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 5452
: TYPE: DNA
: ORGANISM: YEBNA
: US-09-130-114-1

```

Query Match	2.58;	Score 87.2;	DB 2;	Length 5452;
Post Local Classification	47.08;	Score 87.2;	DB 2;	Length 5452;

Matches	378;	Conservative	0;	Mismatches	403;	Indels	9;	Gaps	4;
<p> </p>									

RESULT 15
US-08-910-647-1/c
; Sequence 1, Application US/08910647
; Patent No. 6251433

```

1 GENERAL INFORMATION:
2 APPLICANT: Zuckerman et al.
3 TITLE OF INVENTION: Compositions and Methods for
4 TITLE OF INVENTION: Polynucleotide Delivery
5 NUMBER OF SEQUENCES: 4
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Chiron Corporation
8 STREET: 4560 Horton Street
9 CITY: Emeryville
10 STATE: California
11 COUNTRY: U.S.A.
12 ZIP: 94608-2916
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION NUMBER:
19 APPLICATION NUMBER: US/08/910,647
20 FILING DATE:
21 CLASSIFICATION: 514
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Fujita, Sharon M.
24 REGISTRATION NUMBER: 38,459
25 REFERENCE/DOCKET NUMBER: 1218.002
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (510) 923-2706
28 TELEFAX: (510) 655-3542
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 9600 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 US-08-910-647-1

```

Db 1566 CTTTTCACAGTCACAGACTCTCCCTGCTCTTTCACAGACTTCCCTCCGCTCT 1507
QY 265 accgcccccaactcccttcccaacttcaatgacacgcgcacacgcctccctctcaac 324
Db 1506 TTTCAAGTCTTCACGCGGCGGCTCCTACACTCTTCAGACCGGCGCTCCACTACCTCC 1447
QY 325 tccctatgcccgcgcgtgccttccccaactccctccgcctctctctcttcttcccaaa 384
Db 1446 TCGACCCCGGCTCCACTGCTCTGACCCCGGCTCCACACTCTGCTCTCTGCCCCCTCC 1387
QY 385 ----ccacttccctcccaactcccaactcaatagctagaacacaaagacgcgcgcgcg 440
Db 1386 TGTCTCTCTGCT 1327
QY 441 ccaatggtcgcgaagc 500
Db 1326 CCTCTCTGCT 1268
QY 501 cggcggaacctctcgc 560
Db 1267 CTCTGCCCCCTCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1208
QY 561 cgtctcgc 620
Db 1207 CTCTGCCCCCTCTGCT 1150
QY 621 cgttcagacacacgc 680
Db 1149 TGCCCCCT 1090
QY 681 cgcagcg 740
Db 1089 TGCT 1030
QY 741 acaatcattcggaagc 800
Db 1029 TCT 970
QY 801 gcgagatataccacccctgcaaccccgcgctcgctgctctgagcttcgcgcctcgct 860
Db 969 CCT 910
QY 861 cccctctcgc 920
Db 909 TCT 850
QY 921 tccctgacgc 978
Db 849 TCT 790
QY 979 gcttcgc 988
Db 789 CCT 780

Search completed: April 23, 2002, 06:03:43
Job time: 54289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2002, 14:34:04 ; Search time 2858.62 Seconds
(without alignments)
13194.367 Million cell updates/sec

Title: US-09-265-585c-95

Perfect score: 3510
1 ctgctagctcagcctactca.....tagccatcaagcaagtatla 3510

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estro:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gp_est1:*
11: gp_est2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_huv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_tod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
c 1	498.4	14.2	715	10	AU093320 AU093320
c 2	466.2	13.3	586	10	BE359444 DG1_53.C1
c 3	378.6	10.8	665	10	BE366082 P11_30.C1
c 4	201.8	5.7	528	11	BG154957 sab28b08.
c 5	194.6	5.5	431	10	BE209394
c 6	193.2	5.5	504	10	AM004485
c 7	175	5.0	1015	11	AM004485 701931804
c 8	170.2	4.8	1223	11	B1416573
c 9	165	4.7	1197	11	B1416573 hasp001xg
c 10	164	4.7	484	11	BE265162 HV_CEA001
c 11	163.4	4.7	728	11	B1311576 EST531332
c 12	162.4	4.6	1299	11	BE267681 HV_CEA001

c 13	161.8	4.6	1188	10	BE455091
c 14	161.6	4.6	931	11	B1416575
c 15	161	4.6	1050	11	BE267685
c 16	159.8	4.6	893	13	AZ199082
c 17	155.8	4.4	937	11	B1416476
c 18	155.4	4.4	1039	11	BE259783
c 19	155	4.4	1224	11	B1416537
c 20	154.8	4.4	1080	10	BE455154
c 21	154.6	4.4	1042	11	B1416535
c 22	152	4.3	1135	11	BE256506
c 23	152	4.3	1161	11	BE256588
c 24	150.4	4.3	1195	11	B1416506
c 25	149.4	4.3	1088	11	BE256580
c 26	149.2	4.3	1277	11	BG447302
c 27	147.8	4.2	1107	11	B1416477
c 28	147.8	4.2	1342	10	BE455093
c 29	147	4.2	924	10	BE455162
c 30	146	4.2	1035	11	BE254665
c 31	145.4	4.1	1006	10	BE034822
c 32	145.4	4.1	1116	11	BE256617
c 33	145.2	4.1	1248	10	BE455160
c 34	144.6	4.1	1325	11	BE256543
c 35	143.8	4.1	563	11	BE588097
c 36	143.6	4.1	952	11	BG441341
c 37	143	4.1	883	13	AZ186646
c 38	143	4.1	1028	11	BG809907
c 39	142.6	4.1	700	10	AL506960
c 40	142.6	4.1	1080	13	AO893056
c 41	141.8	4.0	1166	11	BE256751
c 42	141.6	4.0	999	11	B1416539
c 43	141.4	4.0	1195	11	BE256755
c 44	141	4.0	1328	10	BE455227
c 45	140.4	4.0	551	10	A1770444

ALIGNMENTS

RESULT 1
LOCUS AU093320/c 715 bp mRNA EST 21-JUN-2000
DEFINITION AU093320 Rice callus Oryza sativa cDNA clone C61354, mRNA sequence.
ACCESSION AU093320
VERSION AU093320.1 GI:8594782
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 715)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2000)
Unpublished (2000)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abdr.affrc.go.jp, URL: http://irgp.dna.affrc.go.jp/PROJECT/"RGP".
C61354_82.

FEATURES
source location/Qualifiers

1..715
/organism="Oryza sativa"
/strain="Cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="C61354"
/clone_1bp="Rice callus"
/note="Vector: pBluescript II SK+; Site 1: SalI; Site 2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site

BASE COUNT 109 a 261 c 227 g 116 t 2 others
 ORIGIN

Query Match 14.2%; Score 498.4; DB 10; Length 715;
 Best Local Similarity 87.1%; Pred. No. 5e-65;
 Matches 581; Conservative 0; Mismatches 82; Indels 4; Gaps 3;

QY 1132 caaagagcgaagagagagcagcgcggaagcagcgcgagagagagagcctcactgct 1191
 |||||
 DB 715 CAAGAGACGGAGAGAGACGCGGGAAGCAGCGAGAGAGGCGCTCCACTGCT 656
 QY 1192 gacgcctcgtcagtcgagcgcgagcgcgtgaacgcggaacacctgagcagcagcaca 1251
 |||||
 DB 655 GACGCTGCTCTCCAGTGCAGCGCGAGAGTGGTAACCGCGACACCTGACGAGGCGCACCG 596
 QY 1252 gacgctgctgagatcgcgagagctagacgcgcgttcgcacacctcgcagcagcgtgagc 1311
 |||||
 DB 595 CGGCTGCTGGAGATCGCGAGAGCTTGCCACGCCGTTCCGACATCCAGCAGAGCGCTCGC 536
 QY 1312 cgcctactcgcgagagcactgctgcgcgcgtcgcacgtcctcgtcgtgctgagc 1371
 |||||
 DB 535 CGCTACTTGGCGAGGCGCATGTCGGGCGGCTGTGAGCTCGTGGGCTGTACGC 476
 QY 1372 ggc 1431
 |||||
 DB 475 GCGGCTCCCAACCGCGTCCCGCGGCGGCGGCGGCTCTCCAGGCGGCGGCTCGCGGCGGCTT 416
 QY 1432 ccaggtgttcaacgcgcacacagcccttcgtaagttctgcacctcaaccgcaacagagc 1491
 |||||
 DB 415 CCAGGTGTCAACGCGCATCAGCCCGTTCGAGATTCTCCACTCAGCGCGCAACCGACGC 356
 QY 1492 catcagaagagcgttcgaagcgcgagagagcgcgtgcacatcatcagctgagcactgaga 1551
 |||||
 DB 355 GATCCAGAGAGCGCTTCCAGAGAGAGAGAGAGGTCACATCATGACCTGGACATCATGCA 296
 QY 1552 ggggctcagctg-cgcgggctctctccacatctctcctccgcgcgcgcgcgcgcgcgcgc 1610
 |||||
 DB 295 GGGGCTCAATGGCGCGGGGCTGTCCACATCTGGGGGTGAGAGCGGGGGGCGCGCGCA 236
 QY 1611 gggtagagctcaacgcgcctcgcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1670
 |||||
 DB 235 GGGTAGAGCTGACCGGGCTG8GGGCGCTCCATGAGAGGCGTGAAGCGGAGCAAGAGCG 176
 QY 1671 tctccgagcttcgc 1730
 |||||
 DB 175 TATCGACTTCGCGGACACGCTGGGATTGCCATTCCAGTTCTCGCGGTGCTGACAGAG 116
 QY 1731 ccgagcaatgttgaccgcgag-aagctaggggtcagcagc--ggagagccgttcgcgcgcgc 1787
 |||||
 DB 115 CCGGATCTTGAACCGGAGAAAGCTAGCGCTNACGCCGCGAGAGCGCGCTGGCGTCC 56
 QY 1788 actgctc 1794
 |||||
 DB 55 ACTGGCT 49

RESULT 2
 LOCUS BE359444 586 bp mRNA EST 20-JUL-2000
 BE359444
 DEFINITION DGI_53_c11_b2_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA
 BE359444
 ACCESSION BE359444
 VERSION BE359444.1 GI:9301001
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 586)
 Cordonnier-Pratt M.-M., Gingle A., Marsala C., Sudman M. and Pratt

TITLE 'L.H.
 JOURNAL An EST database from Sorghum: dark-grown seedlings
 Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 525
 POLYA-No.

FEATURES

source

1..586
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Dark Grown 1 (DGI)"
 /note="Organ: 5-day-old dark-grown seedlings; Vector:
 lambda Zap, Site-1: XhoI, Site-2: EcoRI. The library was
 made from poly-A RNA in the cloning vector lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 87 a 237 c 195 g 67 t
 ORIGIN

Query Match 13.3%; Score 466.2; DB 10; Length 586;
 Best Local Similarity 90.5%; Pred. No. 3.1e-60;
 Matches 524; Conservative 0; Mismatches 43; Indels 12; Gaps 2;

QY 907 gaagagagatgtctccgcagcagc---cgctcgcgcgcgcgcctccgcgcgcgcgcgcgcgc 963
 |||||
 DB 8 GCAGACAGATGCTTCTCTGTCAGAGGCGTGGGCTCCGGCCCGCGGGGCTGACGCT 67
 QY 964 cctctcccccgcacacgcgtctccgcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1014
 |||||
 DB 68 CCTCTCCCTCCACGCGTTCAAGACAGACGCGCACAGCGCGCGCTCTCTCTCTCC 127
 QY 1015 gtgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1074
 |||||
 DB 128 TCCGAGACGACAGCAGCAGAGAGAGCCGATCCGCGCTGCGAGTGCACCAAGGCCGAGC 187
 QY 1075 cgcgagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1134
 |||||
 DB 188 CCGGAGAGAGACGCGCGCTGACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 247
 QY 1135 ggaagcgaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1194
 |||||
 DB 248 GAGCGGAG 307
 QY 1195 gctgctgtcgcagctgc 1254
 |||||
 DB 308 GCTGCTGTGTCAGTGTGCGGAGAGCGGTGAACGGGACCACTCGAGCAGCGCACAGAC 367
 QY 1255 gctgctgagatcgc 1314
 |||||
 DB 368 GCTGCTGAGATCGCGGAGACTAGCCAGCCGCTTCGCGACCTCGAGCAGCGCGCGTGGCGCGC 427
 QY 1315 ctacttcgc 1374
 |||||
 DB 428 CTACTTCGGGAGGCGCATGTGCGCGCGCTGTCAGTGTGCTGAGGCGCTGTACGCGCC 487
 QY 1375 gctgc 1434
 |||||
 DB 488 GCTGCGCGCGGAGACCCCGCGCGCGCGCTTCACAGCGCGCGCTGCGCGCGGCTTCA 547
 QY 1435 ggtgttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1473
 |||||
 DB 548 GGTGTTCAGAGGCGATCAGCCCTTGTCTCAAGTTCTCACA 586

RESULT 3
 BE366082 665 bp mRNA EST 20-JUL-2000
 LOCUS P11_30_G11_g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 DEFINITION mRNA sequence.
 ACCESSION BE366082
 VERSION BE366082.1 GI:9307639
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 665)
 AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 ,L.H.
 TITLE An EST database from Sorghum: pathogen-induced plants
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordomier-Pratt MM
 Department of Botany
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@atluga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolYTmix
 High quality sequence start: 16
 High quality sequence stop: 665
 POLYA-No.

FEATURES

source

1..665
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /note="Organ: Anthracnose-induced 1 (P11)"
 Vector: PBluescript II from Lambda Zap II; Site.1: XhoI;
 Site.2: EcoRI; Two-week-old sorghum plants (BX 623
 cultivar) were infected with pathogen (isolate PMW421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT 123 a 197 c 187 g 158 t
 ORIGIN

Query Match 10.8%; Score 378.6; DB 10; Length 665;
 Best Local Similarity 96.5%; Pred. No. 3.5e-47;
 Matches 387; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2756 ctccctggcgcgtctgtgagccatccactactactcggcgtctgtgactcgtgga 2815
 |||||||
 Db 1 CTTCCTGGCCGCTGCTGTGAGGCGCATCTACTACTGCGCCTGTTCAGCTCGCTGG 60
 |||||||
 QY 2816 cgcgagctacgcgcgagacagcccgagcgacgctgtgagcaagcagctgtcgcgcg 2875
 |||||||
 Db 61 CGGAGAGTACGGGAGAGAGAGCCCGAGCGGACGTCGTGACACAGAGCTGTGTCGG 120
 |||||||
 QY 2876 gggagatccgaagctgtgtgcctgtggcgccgagcccgacacgcgcgaagtcagg 2935
 |||||||

Db 121 GGAGATCCGACAGCTGCGCCGCGCCGCCGCCACCGGCGACGTCAAGTCCG 180
 QY 2936 cagctgagcgaagaagctgagcagctcgggttcgcgcgcgcctcgtcgcgcgcgcgc 2995
 |||||||
 Db 181 CACTGTGGCGGAGAACACTGCGGAGTCCGGGTTCCCGCGCCTGCTGCGCGGAGCGC 240
 |||||||
 QY 2996 cgcgcgcgcgcgcgtccctcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3055
 |||||||
 Db 241 CGCCGCGGAGAGGCTGCTGCTGCTGCGAATGTTCCCTCGACGGGTACAGCTGTGTA 300
 |||||||
 QY 3056 ggaagacgc 3115
 |||||||
 Db 301 GGAGAACGCGCGCGCTTACCTCGCGGAGGACCTCTCTCTCCACCGCGCTCGCATG 360
 |||||||
 QY 3116 ggc 3156
 |||||||
 Db 361 GCGCCCATCCAGATCCCGCGCGCTGATGAGAGCTGT 401
 |||||||

RESULT 4

BG154957

LOCUS

DEFINITION

BG154957 528 bp mRNA EST 06-FEB-2001
 sab28b08.y1 Gm-cl026 Glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-cl026-2775 5' similar to TR:096304 Q96304 SCARECROW.; mRNA
 sequence.

ACCESSION

BG154957

VERSION

BG154957.1

KEYWORDS

EST.

SOURCE

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE

AUTHORS

1 (bases 1 to 528)

Shoemaker,R., Kelm,P., Vodka,L., Expelding,J., Coryell,V., Khanna,
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 for further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 455.

FEATURES

source

1..528
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl026-2775"
 /clone_lib="Gm-cl026"
 /tissue_type="Senescing leaves, mature plants, greenhouse
 grown."
 /lab_host="DH10B"
 /note="Vector: pT73-Pac (Pharmacia); site_1: EcoRI;
 site_2: HindIII; This cDNA library was constructed from
 mRNA isolated from senescing leave tissue of mature
 greenhouse grown plants. Complementary DNA was synthesized
 from mRNA using a 3' anchored poly(dT) primer. EcoRI
 adapters were ligated to the blunt-ended cDNA fragments
 followed by digestion with EcoRI and HindIII. The cDNA
 fragments were directionally cloned into the EcoRI-HindIII

Df	361	ACCTTGGTTGTCACACTGCATCAGCTTGGAACC	394
RESULT	6		
Locus	AM004485	504 bp	mRNA EST
DEFINITION	701931804 A. thaliana, mixed source Arabidopsis thaliana cDNA clone		08-SEP-1999
ACCESSION	701931804, mRNA sequence.		
VERSION	AM004485		
KEYWORDS	AM004485.1 GI:5851514		
SOURCE	EST.		
ORGANISM	thale cress.		
AUTHORS	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (bases 1 to 504)		
REFERENCE	Chen,J., Montyama,M., Chan,E., Mooney,M., Carreon,B., Gilliland,D., Wang,K., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzaska,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutoula,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Caprio,T., Polcify,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L., Turner,C., Kikorian,S., Elder,L. and Hanson,D. Arabidopsis thaliana Gene Expression Microarray Unpublished (1999) Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 463 World Parkway Circle, St. Louis, MO 63134, USA Tel.: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.		
JOURNAL COMMENT			
TITLE			
FEATURES			
source	Location/Qualifiers		
	1..504	/organism="Arabidopsis thaliana"	
		/db_xref="taxon:3702"	
		/clone="701931804"	
		/clone_1ib="A. thaliana, mixed source"	
		/note="This sequence was obtained from a clone generated with a PCR product of the target gene."	
BASE COUNT	137 a 93 c 130 g 144 t		
ORIGIN			
Query Match	5.5%; Score 193.2; DB 10; Length 504;		
Best Local Similarity	66.1% ; Pred. No. 1.5e-19;		
Matches	279; Conservative 0; Mismatches 14; Indels 0; Gaps 0;		
OY	2700 agcgctgcggcccaaggttgatgaacatgltgagaacgaactgaagccaactcggctccctc	2759	
Dd	73 AGATTAGCTCCATAAAAGTTGTGTGACACTAGTAGCAAGAATTGAGCCACGCTGTTCTTWC	132	
OY	2760 ctgcgcgcttcctgtagaacgccatcacactactacttggcgcttgttcgactcgtgtgaagg	2819	
Dd	133 TTAGGAAATTTGTGTAAGAACCATTAATTACTACTCTTGACACTCTTTGACCTCGAGACA	192	
OY	2820 agctacggcgaaagacaagcccagcgagcgacgtcgtgagcaacgaactcgtcgcggag	2879	
Dd	193 AGCTACGGCGGAAGAGTAGAAGAGACATGTCGTGGAAACACACTATTATTCGAAGAG	252	
OY	2880 atccgcaaatcgtcgtgcgctggcgggcgccgcgcacaccgcgacgtcaagtccgcaag	2939	
Dd	253 ATACGGAATGTTTTAGCGTTGTGAGGACCATCGAAGACGGGTGAAGTTGAGAGC	312	
OY	2940 tggcgcggaagactcgtgcgcagtcgagttccgcgcgcctgcgtccgcgcgcgcgcgcgcg	2999	
Dd	313 TGGAGGGAAGAAAATCACACAAATGTGGGTTTAAAGGTATATATCTTTAGCTGGAATCCAGCT	372	
OY	3000 gccgcgcgcctcctctgctgcgcatagttccccctcgaaggttaacgctgtgtgagag	3059	
Dd	373 ACACAAGGACACTCTACTGTGTGGAAATGTTTCCCTTGGAGTGTTAACACTTGTGTTATCAT	432	
OY	3060 aaagcgcgctgaagctcgggtgtgaaggaacctctgctgtccaccgcgctcgcctcgcgc	3119	

[illegible]

[illegible]

RESULT	9
BI416470/c	
LOCUS	BI416470
DEFINITION	1197 bp mRNA EST 15-AUG-2001 hasp001xa01f Heterobasidium annosum - Scots pine infection stage

ACCESSION	BI416470	
VERSION	BI416470.1	GI:15187493
KEYWORDS	EST.	

SOURCE Heterodactylodon amnosum/Pinus sylvestris mixed EST library.
ORGANISM Heterodactylodon amnosum/Pinus sylvestris mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Astieghu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the

Journal Unpublished (2001)
Comment Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture Box 7026 S-750 07 Uppsala

FEATURES
 www.cmu.edu
 Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Astiegbu@nykopat.slu.se
 Seq primer: T7 primer.
 location/Qualifiers

```

/organism="Heterobasidion annosum/Pinus sylvestris mixe
Est library"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infecti

```

`/dex_stage="Seedling roots of scots pine were infected
6 days with H. annosum"
/note=Vector: pt-Adv; Site_1: EcoRI. The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of`

BASE COUNT	20 a	28 c	1036 g	28 t	85 others
ORIGIN					
Query Match		4.7%	Score 165;	DB 11;	Length 1197;

Matches	462;	Conservative	0;	Mismatches	545;	Indels	2;	Gaps	
---------	------	--------------	----	------------	------	--------	----	------	--

[illegible]

Dy 461 ggcgacagactcctccggcgccgcgcgaatcaaggcgacattctcgagt 520
 || | | | | | | | | | |
Bd 822 CNNCCCCCCCCCCCNCCNCCCNCNCCNCCNCCNCCNCCNCCNCCNCC 763

581 ctgcgcgcgtgccaccagctccaccagctgcgcgcggttcagcacccgcgcg 640

[illegible]

D_b 582 CCCCCCCCNCNCCGCCACCAACCACCAACCCCCCNCCNCCCNCCCCCCCCAACCC 523
Q_y 761 ggcgcgcggtctccatcagcgacgtcatcacaaagtccgcgagatcatccaccctgc 820

D6	522 CCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	463
Q7	821 aaaccgcgctcgttcgtctctgagctcgctccctcctcagccgacgg	880
D6	462 CCCCCCCCCCCCCCCCCCNCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCC	403

0Y 941 ggcgcctccgcgcggcgtgacgtccctcccccgcacgcgcgttcgcgagcaagcgcgcgcac 1000

DQ 284 GCGATCCACGCGCTGCAGCAGCACAGCAGGAACGCGATCCGGCGCAATCG 1060
|||
Dd 1001 GCGATCCACGCGCTGCAGCAGCACAGCAGGAACGCGATCCGGCGCAATCG 1060
|||||
Dc 284 NCCNNCCNCCCNCNAAACCCCNCCCCCNCOTCCCCGCCNCCCCTNCCCCTCCMCC 225

Db 224 CCGTCCNNCCGGCCNTNCCAGGCCCCCGCGGCCNAGCCCCCCC 176

LOCUS	BF265162	484 bp	mRNA	EST	09-MAR-2001
-------	----------	--------	------	-----	-------------

DEFINITION	HV.CEA0011H03f Hordeum vulgare seedling green leaf EST library HYCDNA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone HV.CEA0011H03f, mRNA sequence.
ACCESSION	BF265162
VERSION	BF265162.2 GI:13261932
KEYWORDS	EST.
SOURCE	barley.
ORGANISM	Hordeum vulgare
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum. 1 (bases 1 to 484) Wing,R., Close,T.J., Kleinhoofs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo ,T., Sakri,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and Wood,T. Development of a genetically and physically anchored EST resource for barley genomics Unpublished (2000) On Nov 17, 2000 this sequence version replaced gi:11196156. Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: AATTAACCCCTCACTAAAGG High quality sequence stop: 431. Location/Qualifiers 1..484 /organism="Hordeum vulgare" /cultivar="CI16155 (M1a13)" /db_xref="taxon:4513" /clone="HV_CEA0011H03f" /clone_lib="Hordeum vulgare seedling green leaf EST library HYCDNA0004 (Erysiphe infected & control)" /lssue_type="seedling green leaf" /lab_host="TJC121" /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley/To order a clone see http://www.genome.clemson.edu/orders"
FEATURES	
Source	
BASE COUNT	101 a 137 c 134 g 112 t
ORIGIN	
Query Match	4.7%; Score 164; DB 11; Length 484;
Best Local Similarity	69.0%; Pred. No. 3.3e-15;
Matches	254; Conservative 0; Mismatches 110; Indels 4; Gaps 2;
Oy	2862 cagctgcgttcgcgggaagatccgaagtgctcgtgcggttgcggcgccgcgcacgcgcg 2921
Db	27 CAGCGTGCGGGGAGGAGCATCCGCATGTCTTGCGTGTGGCGGCCGCCGACACC6GG 86
Oy	2922 gacgtcaagttccgaagtgcgcgaagaagcttgcgcagtcgcagggttccgcgcgcctcg 2981
Db	87 GACATCAAGTTCCGAACACTGGCGGGAGAAGCTTCGGCGCAGTGGGGTTCCGTGCGGCTCG 146
Oy	2982 ctccgcgcgcagcgcgcgcgcgcgcgcgcgcgtccctcgtccttcgcgcacatgttccctccgcagcg 3041
Db	147 CTGGCTGGCACGCGCGCAGCGCGCATCCCTCTGCTGTCGCGCGCATGTCCCACTCGGATGCG 206
Oy	3042 tacaaagctgcggtagaagaagcgcgctcgtgaag-ctcgggtgcgaagagaccctcgtcgtc 3100
Db	207 TACAATTATACGTCGTGACAAAATGCGCTCTTAAAAAGCCTTGATGGAAGATCAGTGGCTTGGT 266
Oy	3101 cacccgcgtgcgcgttcgcgcgcccatccaagtgc---gcgcgtgcgcgttatagaacctcgt 3157
Db	267 TCCTGCTGGAGACTTCGCCCCCAATTTCAGGCGTCAATGTCGATGAATAATTACCTAGAGCTA 326
Oy	3158 cctgctcctcgtcgttcgttcgaagagcgccgacctccactgttttgcattcgttaagctgcgcg 3217

Db	327	AACCATGCCATTCCCTATTGGCTGCGGCTGTCTCTGTTTCCGTACTTAAAGTGCCTCG	386
QY	3218	gtttggtc 3225	
Db	387	GTGGCATC 394	
RESULT	11		
LOCUS	BI111576	728 bp	mRNA EST 20-JUL-2001
DEFINITION	EST5131366 GESD Medicago truncatula cDNA clone pGESD13C12 5' end,		
ACCESSION	BI111576		mRNA sequence.
VERSION	BI111576.1	GI:14985903	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;		
	Medicago.		
REFERENCE	1 (bases 1 to 728)		
AUTHORS	Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho		
TITLE	,J. and Fraser,C.M.		
JOURNAL	ESTs from developing reproductive tissues of Medicago truncatula		
COMMENT	Unpublished (2001)		
	Contact: Michael A. Grusak		
	USDA/ARS Children's Nutrition Research Center		
	Baylor College of Medicine		
	1100 Bates Street, Houston, TX 77030-2600, USA		
	Tel: 713-798-7044		
	Fax: 713-798-7078		
	Email: mgrusak@bcm.tmc.edu		
	B398431e		
FEATURES			
source	TIGR sequence name: MTPAV187K		
	More information is available at: www.medicago.org		
	Seq primer: SKmod (CTA gAA CTA gTg gAT CC).		
	Location/Qualifiers		
	1..728		
	/organism="Medicago truncatula"		
	/cultivar="A17"		
	/db_xref="taxon:3880"		
	/clone="pGESD13C12"		
	/clone_lib="GESD"		
	/tissue_type="Immature seeds"		
	/dev_stage="Immature seeds, 11 to 19 days after		
	pollination"		
	/note="Vector: pluescript SK-; Site_1: EcoRI; Site_2:		
	XhoI; Immature seeds, collected from pods ranging in age		
	from 11 to 19 days after pollination, were harvested from		
	greenhouse-grown plants. Seeds were removed and		
	separated from pod walls and were immediately frozen in		
	liquid nitrogen. Seeds throughout the age range were		
	pooled for mRNA extraction. cDNA was prepared from polyA+		
	enriched RNA. The cDNA was directionally ligated into		
	the unzip xr vector from stragene and packaged using		
	gigaquick III gold packaging extracts. Plasmids containing		
	cDNA inserts were excised from the recombinant lambda-zap		
	phage using Ex-assist helper phage and propagated in		
	XLOIR cells."		
BASE COUNT	204 a 123 c 168 g 233 t		
ORIGIN			
Query Match	4.7%; Score 163.4; DB 11; Length 728;		
Best Local Similarity	64.3%; Pred. No. 4.1e-15;		
Matches 245; Conservative	0; Mismatches 136; Indels 0; Gaps 0;		
2741	gagcaaccggagctctccctcgcgcgcgttgtagagacatccacactactacgcgcct 2800		
Db	1 GAGCAATCGACAGGCTCATTTCTTGCGAAGGTTTGCAAGCAATACATTCTCTACGACATT 60		
2801	gttcgaccctcgtctgacgagagctacgcgagaggaacagcccccagcgacgctctgtagaca 2860		

[illegible]

RESULT	12			
LOCUS	BF267681			
DEFINITION	BF267681	1299 bp	mRNA	EST
	HV.CE80018J13f Hordeum vulgare seedling green leaf EST library			
	HVDNA00004 (Erysiphe infected & control) Hordeum vulgare cDNA clone			
	HV.CE80018J13f, mRNA sequence.			

ACCESSION	BF267681
VERSION	BF267681.1
KEYWORDS	GI:11198676
EST.	

SOURCE	ORGANISM
barley.	<i>Hordeum vulgare</i>
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.
REFERENCE	1 (bases 1 to 1299)
AUTHORS	Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and Wood,T.

TITLE	Development of a genetically and physically anchored EST resource for barley genomics
JOURNAL	Unpublished (2000)
COMMENT	Contact: Wing RA

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twining@clemson.edu
Seq primer: AATTACCTCTACTAAGG
High quality sequence start: 321
High quality sequence stop: 1283.

FEATURES	Location/Qualifiers
source	1. .1299

```

/organism="Hordeum vulgare"
/cultivar="C16155 (Ma13)"
/db_xref="taxon:4513"
/clone="HY_CEA0018J13f"
/clone_id="Hordeum vulgare seedling green leaf EST
library HVCDNA0004 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"

```

```

/lab_host="TJCI21"
/notes=vector: lambdaZAP: Site_1: EcoRI: Site_2: XhoI: For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/To
Order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT      0 a      827 c      82 g      0 t      390 others
ORIGIN

```

Query Match	4.68;	Score 162.4;	DB 11;	Length 1299;
Best Local Similarity	42.98;	Pred. No. 6.1e-15;		
Matches 441; Conservative	0;	Mismatches 583;	Indels 4;	Gaps 1;

[illegible]

